

**TRANSACTIONS**  
**OF THE**  
**AMERICAN MATHEMATICAL SOCIETY**

**EDITED BY**

**WILLIAM C. GRAUSTEIN**

**EINAR HILLE**

**C. C. MACDUFFEE**

**WITH THE COÖPERATION OF**

**A. A. ALBERT  
T. H. HILDEBRANDT  
MARSTON MORSE  
H. P. ROBERTSON  
GABOR SZEGÖ**

**ERIC T. BELL  
E. P. LANE  
OYSTEIN ORE  
M. H. STONE  
G. T. WHYBURN**

**JESSE DOUGLAS  
R. E. LANGER  
H. L. RIETZ  
J. L. SYNGE  
OSCAR ZARISKI**

**VOLUME 45  
JANUARY TO JUNE, 1939**

**PUBLISHED BY THE SOCIETY  
MENASHA, WIS., AND NEW YORK  
1939**

**BOSTON UNIVERSITY  
COLLEGE OF LIBERAL ARTS  
LIBRARY**

*Charles C. Carter Library Fund*  
*Bound Sept. 1939*

Composed, Printed and Bound by  
*The College Press*  
George Banta Publishing Company  
Menasha, Wisconsin

61026

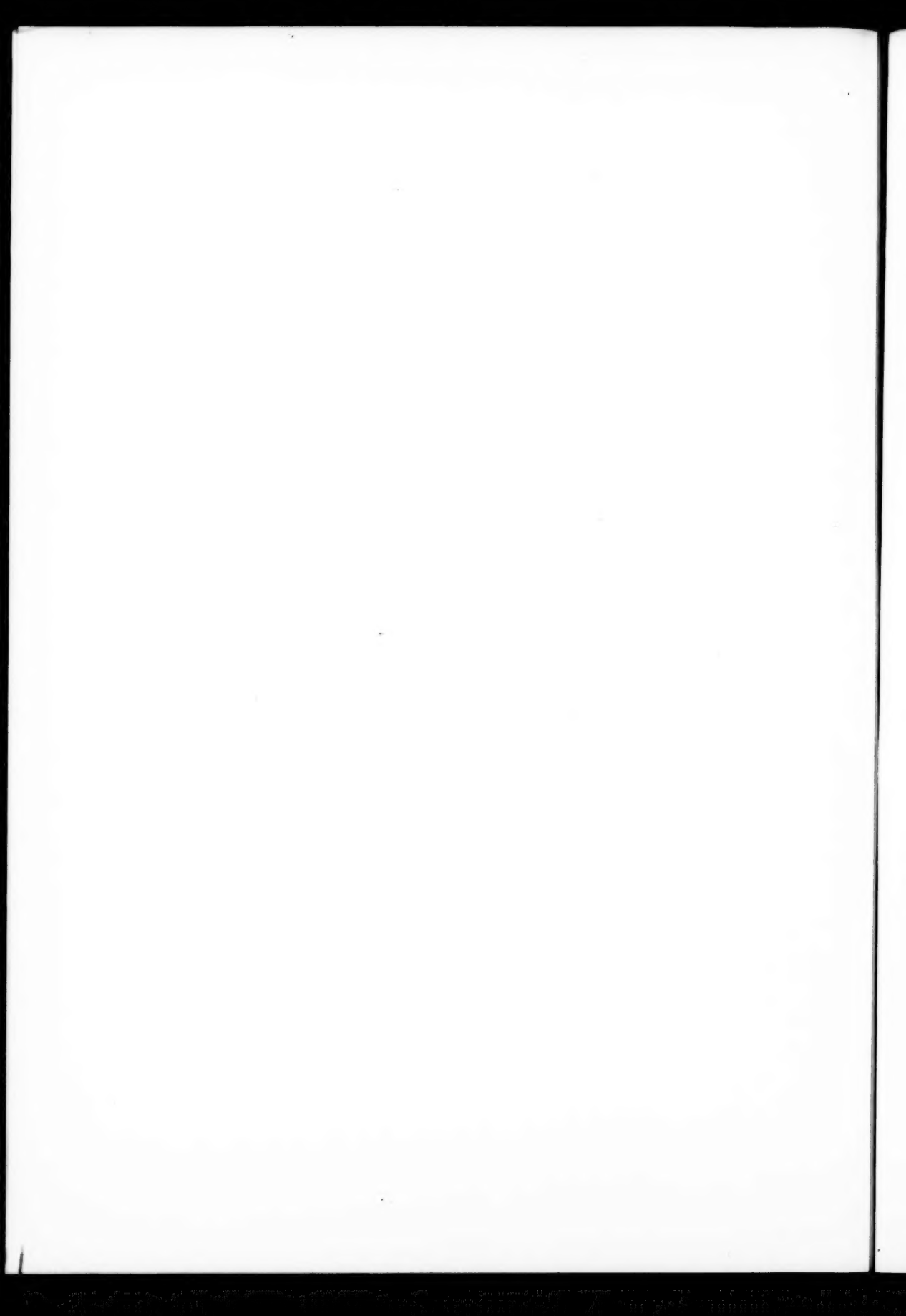
QA  
1  
F001  
v.45

22

# TABLE OF CONTENTS

## VOLUME 45, JANUARY TO JUNE, 1939

ADAMS, C. R., and CLARKSON, J. A. The type of certain Borel sets in several Banach spaces.....	322
BOAS, R. P., and WIDDER, D. V. The iterated Stieltjes transform.....	1
CALKIN, J. W. Abstract symmetric boundary conditions.....	369
CLARKSON, J. A., and ADAMS, C. R. The type of certain Borel sets in several Banach spaces.....	322
COXETER, H. S. M. The abstract groups $G^{m,n,p}$ .....	73
DILWORTH, R. P., and WARD, M. Residuated lattices.....	335
FIALKOW, A. Conformal geodesics.....	443
GROVE, V. G. Geometry of a surface in the neighborhood of a spine.....	256
LORCH, E. R. On a calculus of operators in reflexive vector spaces.....	217
MC SHANE, E. J. Some existence theorems in the calculus of variations.	
III. Existence theorems for nonregular problems.....	151
MC SHANE, E. J. Some existence theorems in the calculus of variations.	
IV. Isoperimetric problems in non-parametric form.....	173
MC SHANE, E. J. Some existence theorems in the calculus of variations.	
V. The isoperimetric problem in parametric form.....	197
MAKER, P. T. Conditions on $u(x, y)$ and $v(x, y)$ necessary and sufficient for the regularity of $u+iv$ .....	265
MARDEN, M. Kakeya's problem on the zeros of the derivative of a polynomial.....	355
MURRAY, F. J. Bilinear transformations in Hilbert space.....	474
SOKOLNIKOFF, E. S., and SOKOLNIKOFF, I. S. Thermal stresses in elastic plates.....	235
SOKOLNIKOFF, I. S., and SOKOLNIKOFF, E. S. Thermal stresses in elastic plates.....	235
STRODT, W. C. Irreducible systems of algebraic differential equations.....	276
WARD, M., and DILWORTH, R. P. Residuated lattices.....	335
WIDDER, D. V., and BOAS, R. P. The iterated Stieltjes transform.....	1
ZUCKERMAN, H. S. On the coefficients of certain modular forms belonging to subgroups of the modular group.....	298



# THE ITERATED STIELTJES TRANSFORM\*

BY

R. P. BOAS, JR. AND D. V. WIDDER

## INTRODUCTION

This paper is concerned with the iterate of the Stieltjes transform

$$(1) \quad f(x) = \int_0^{\infty} \frac{d\alpha(t)}{x+t},$$

which in turn is the iterate of the Laplace transform

$$(2) \quad f(x) = \int_0^{\infty} e^{-xt} d\alpha(t).$$

It is known† that (2) can be inverted by use of the differential operator of infinite order

$$\lim_{k \rightarrow \infty} \frac{(-1)^k}{k!} \left( \frac{k}{t} \right)^{k+1} f^{(k)}(k/t),$$

an operator which annuls the functions

$$(3) \quad f(x) = x^k, \quad k = 0, 1, 2, \dots$$

It is also known‡ that (1) can be inverted by use of the linear differential operator of infinite order

$$\lim_{k \rightarrow \infty} \frac{(-t)^{k-1}}{k!(k-2)!} [t^k f(t)]^{(2k-1)},$$

an operator which annuls the functions (3), where now  $k$  runs through the negative integers as well.

When the transform (1) is iterated, one is led to the transform

$$(4) \quad f(x) = \int_0^{\infty} \frac{du}{x+u} \int_0^{\infty} \frac{d\alpha(t)}{u+t},$$

or, when it is permissible to change the order of integration, to the transform

$$(5) \quad f(x) = \int_0^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t).$$

\* Presented to the Society, October 30, 1937; received by the editors October 1, 1937.

† D. V. Widder, *The inversion of the Laplace integral and the related moment problem*, these Transactions, vol. 36 (1934), p. 107.

‡ See the paper of D. V. Widder cited in §3.

To distinguish between these two cases we refer to (4) as the *iterated Stieltjes transform* and to (5) as the  $S_2$  transform. We show that the existence of the integral (5) implies the existence of the repeated integral (4), but not conversely.

It should be observed that the kernel

$$(x-t)^{-1} \log(x/t)$$

becomes infinite as  $t$  approaches zero. For this reason the integral (5) must be understood to mean

$$\lim_{\epsilon \rightarrow 0, R \rightarrow \infty} \int_{\epsilon}^R \frac{\log(x/t)}{x-t} d\alpha(t).$$

When it is desirable to emphasize that this Cauchy value of the integral is intended, we write it as

$$\int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t).$$

In the first part of the present paper the inversion of the integrals (4) and (5) is discussed. It is found that the inversion operator is again a linear differential operator which annuls the functions (3) for  $k=0, \pm 1, \pm 2, \dots$  and in addition the functions

$$f(x) = x^k \log x, \quad k = 0, \pm 1, \pm 2, \dots$$

For  $k$  an integer greater than unity we define an operator  $H_{k,t}[f]$  by the relation

$$H_{k,t}[f(x)] = \left[ \frac{1}{k!(k-2)!} \right]^2 \{ t^{2k-1} [t^{2k-1} f^{(k-1)}(t)]^{(2k-1)} \}^{(k)}.$$

We are then able to show that if  $\alpha(t)$  is an integral and is of such a nature that (4) or (5) exists, then

$$\lim_{k \rightarrow \infty} H_{k,t}[f(x)] = \alpha'(t)$$

for almost all positive values of  $t$ . If  $\alpha(t)$  is of bounded variation in every finite interval and is of such a nature that (4) or (5) exists, then

$$\frac{1}{2} [\alpha(t+) + \alpha(t-)] = \alpha(0+) + \lim_{k \rightarrow \infty} \int_{0+}^t H_{k,u}[f(x)] du$$

for all positive values of  $t$ .

In the remaining part of the paper necessary and sufficient conditions for

the representation of functions in the forms (4) and (5) are discussed. The most important results are summarized in the following table.

<i>Class of the function <math>\alpha(t)</math></i>	<i>Condition</i>
(A) Non-decreasing	$H_{k,t}[f] \geq 0, (t > 0)$
(B) Of bounded variation on $(0, \infty)$	$\int_0^\infty  H_{k,t}[f]  dt \leq M$
(C) Integral of a function of $L^p, (p > 1)$	$\int_0^\infty  H_{k,t}[f] ^p dt \leq M$
(D) Integral of a function of $L$	$\text{l.i.m.}_{k \rightarrow \infty}^{(1)} H_{k,t}[f]$ exists
(E) Integral of a bounded function	$ H_{k,t}[f]  \leq M, (t > 0)$

$$(6) \quad \begin{aligned} f(x) &= o(x^{-1}), & x &\rightarrow 0+, \\ f(x) &= o(1), & x &\rightarrow \infty. \end{aligned}$$

An entry in the right-hand column of this table indicates that those conditions for an infinite sequence of positive integers  $k$  plus conditions (6) are necessary and sufficient for the representation of  $f(x)$  in the form (4) with  $\alpha(t)$  a function of the class described in the corresponding left-hand column. It is found that an additional condition must be added, except in cases (A) and (C), for representation in the form (5).

The method of proof is such that from the conditions in the right-hand column one must be able to infer that

$$\begin{aligned} f^{(k)}(x) &= o(x^{-k-1}), & x &\rightarrow 0+, \\ f^{(k)}(x) &= o(x^{-k}), & x &\rightarrow \infty, \end{aligned}$$

for all non-negative integers  $k$ . This is done by use of a result of R. P. Boas\* concerning the asymptotic behavior of Euler differential forms.

The proofs of our representation theorems are necessarily complicated by the fact, observed above, that the kernel of equation (5) is not bounded. A bounded auxiliary kernel

$$E(x, t) = Q_z \left[ \frac{\log(x/t)}{x-t} \right] = 2x \int_0^\infty \frac{u^2 du}{(x+u)^3(t+u)^2},$$

where

$$Q[f(x)] = x[x^2 f'(x)]'',$$

is used. Conditions for the representation of  $f(x)$  in the form

$$f(x) = \int_0^\infty E(x, t) d\alpha(t)$$

\* See the reference in §14.

are first obtained, and the transition to the form (4) is then made by use of a Tauberian theorem.

The two inversion formulas, for  $S_2$  transforms, and the representation theorem (C) have been previously announced by D. V. Widder.\*

#### CHAPTER I. PROPERTIES OF THE TRANSFORMS

1. **The  $S_2$  transform.** Let  $\alpha(u)$  be a function, defined on  $(0, \infty)$ , of bounded variation on every interval  $(\epsilon, R)$ ,  $(0 < \epsilon < R < \infty)$ , and normalized by the conditions

$$(1.1) \quad \alpha(0) = 0, \quad \alpha(u) = \frac{1}{2} [\alpha(u+) + \alpha(u-)], \quad u > 0.$$

For  $x > 0$  we consider the limit

$$(1.2) \quad f(x) = \lim_{\epsilon \rightarrow 0+, R \rightarrow \infty} \int_{\epsilon}^R \frac{\log(x/t)}{x-t} d\alpha(t),$$

where  $(\log x - \log t)/(x-t)$  is defined by its limiting value  $1/x$ , for  $t=x$ .

DEFINITION 1.1. For any points  $x > 0$  for which the limit (1.2) exists, defining a function  $f(x)$ ,  $f(x)$  is said to be an  $S_2$  transform, convergent for such points. We write

$$(1.3) \quad f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t).$$

The function  $\alpha(t)$  is called the determining function of  $f(x)$ .

THEOREM 1.1. If the  $S_2$  transform (1.3) converges for some  $x_0 > 0$ , it converges for every  $x > 0$  and converges uniformly in any interval  $a \leq x \leq A$ , where  $0 < a < A < \infty$ .

It is necessary to show that the two integrals

$$(1.4) \quad \int_{2A}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t), \quad \int_{0+}^{a/2} \frac{\log(x/t)}{x-t} d\alpha(t)$$

converge uniformly,  $a \leq x \leq A$ .

To discuss the first integral (1.4), we set

$$\beta(t) = \int_B^t \frac{\log(x_0/u)}{x_0-u} d\alpha(u), \quad \beta(B) = 0,$$

where  $B > 2A$  is a sufficiently large constant. By hypothesis,  $\beta(\infty)$  exists. For  $a \leq x \leq A$  and  $R > B$ ,

\* D. V. Widder, *The iterated Stieltjes transform*, Proceedings of the National Academy of Sciences, vol. 23 (1937), pp. 242-244.

$$\begin{aligned}
\int_B^R \frac{\log(x/t)}{x-t} d\alpha(t) &= \int_B^R \frac{\log x - \log t}{\log x_0 - \log t} \frac{x_0 - t}{x - t} d\beta(t) \\
&= \beta(R) \frac{\log x - \log R}{\log x_0 - \log R} \frac{x_0 - R}{x - R} \\
&\quad - \log \frac{x}{x_0} \int_B^R \beta(t) \frac{x_0 - t}{x - t} \frac{dt}{t(\log x_0 - \log t)^2} \\
&\quad - (x_0 - x) \int_B^R \beta(t) \frac{\log x - \log t}{\log x_0 - \log t} \frac{dt}{(x - t)^2} \\
&= J_1 - J_2 - J_3.
\end{aligned}$$

Now let  $R \rightarrow \infty$ . The term  $J_1$  is  $\beta(R)$  multiplied by a bounded factor which approaches unity, so that  $\lim_{R \rightarrow \infty} J_1 = \beta(\infty)$ , uniformly for  $a \leq x \leq A$ . Since  $\beta(t)$  is bounded, it is easily seen that  $J_2$  is dominated by an integral of the form

$$\int \frac{M dt}{t(C + \log t)^2},$$

where  $M$  and  $C$  are constants, independent of  $x$ , and that  $J_3$  is dominated by an integral of the form

$$\int \frac{M dt}{(C + t)^2}.$$

Thus  $J_2$  and  $J_3$  approach limits as  $R \rightarrow \infty$ , uniformly for  $a \leq x \leq A$ . To treat the other integral (1.4), we consider

$$(1.5) \quad \lim_{\epsilon \rightarrow 0+} \int_{\epsilon}^{a/2} \frac{\log(x/t)}{x-t} d\alpha(t),$$

and set  $x = y^{-1}$ ,  $t = u^{-1}$ , and  $\epsilon = R^{-1}$ ; the limit (1.5) becomes

$$\lim_{R \rightarrow \infty} y \int_{2/a}^R \frac{\log(y/u)}{y-u} d\gamma(u), \quad \gamma(u) = \int_{2/a}^u (-v) d\alpha(v^{-1}),$$

which is a limit of the form already discussed.

We have defined the  $S_2$  transform only for a real variable  $x$ . If we regard  $x$  as a complex variable in (1.2) and admit all determinations of the logarithmic function, we may still call the function defined by (1.2) an  $S_2$  transform. In this paper we shall not discuss the  $S_2$  transform in the complex domain; but we state here, without proof, some of its properties.

**THEOREM 1.2.** *If an  $S_2$  transform converges for any complex  $x \neq 0$ , with any determination of the logarithm, it converges for every  $x$  not on  $D$ , the positive real axis, with any determination of the logarithm, and for  $x$  on  $D$  if the principal value\* of the logarithm is used. If, in a region  $S$  in which the  $S_2$  transform converges, the determinations of  $\log x$  used form an analytic function, then the  $S_2$  transform represents a function analytic in  $S$ . The analytic function obtained by using the principal value of  $\log x$  and continuing the result analytically has  $x=0$  as a singular point.*

**2. Lemmas on Stieltjes integrals.** We prove the following lemma:

**LEMMA 2.1.** *Let  $f(x)$  be bounded, nonnegative, and monotonic on the (finite) interval  $a \leq x \leq b$ . Let  $\alpha(x)$  have bounded variation on  $a \leq x \leq b$ . Let  $\int_a^b f(x) d\alpha(x)$  exist. Then, according as  $f(x)$  is non-decreasing or non-increasing,*

$$f(b) \text{ l.b. } [\alpha(b) - \alpha(x)] \leq \int_a^b f(x) d\alpha(x) \leq f(b) \text{ u.b. } [\alpha(b) - \alpha(x)],$$

$a \leq x \leq b$   $a \leq x \leq b$

or

$$f(a) \text{ l.b. } [\alpha(x) - \alpha(a)] \leq \int_a^b f(x) d\alpha(x) \leq f(a) \text{ u.b. } [\alpha(x) - \alpha(a)].$$

$a \leq x \leq b$   $a \leq x \leq b$

This lemma will usually appear in the form which it assumes when  $\alpha(x) = \int_a^x \phi(t) d\beta(t)$ , where  $\beta(t)$  is a function of bounded variation on  $a \leq x \leq b$ , and  $\phi(t)$  is a bounded function such that  $\alpha(x)$  is defined. If  $\alpha(x)$  is a Lebesgue integral, the lemma reduces to "Bonnet's theorem,"† since (taking for definiteness the case where  $f(x)$  is non-decreasing)

$$\alpha(x) - \alpha(a) = \int_a^x \alpha'(t) dt,$$

and since the continuous function

$$f(b) \int_a^x \alpha'(t) dt$$

takes on every value between its maximum and minimum, and, in particular, the value

$$\int_a^b f(t) \alpha'(t) dt.$$

\*  $-\pi < \Im[\log x] \leq \pi$ , where the symbol  $\Im$  denotes "imaginary part of."

† See, for example, E. W. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*, vol. 1, 1927, p. 618.

The two cases of the lemma are equivalent, by the substitution  $-y=x$ . We consider the case where  $f(x)$  is non-decreasing. Then

$$\begin{aligned}\int_a^b f(x) d\alpha(x) &= - \int_a^b f(x) d[\alpha(b) - \alpha(x)] \\ &= f(a)[\alpha(b) - \alpha(a)] + \int_a^b [\alpha(b) - \alpha(x)] df(x) \\ &\leq f(a) \text{ u.b. } [\alpha(b) - \alpha(x)] + \text{u.b. } [\alpha(b) - \alpha(x)] \int_a^b df(x) \\ &= f(b) \text{ u.b. } [\alpha(b) - \alpha(x)].\end{aligned}$$

The inequality in the other sense is established similarly.

LEMMA 2.2. Let  $f(x)$  be bounded, nonnegative, and monotonic on the (finite or infinite) interval  $(a, b)$ . Let  $\alpha(x)$  have bounded variation on  $(a+\epsilon, b-\epsilon)^*$  for every (sufficiently small)  $\epsilon > 0$ . Let  $A$  and  $B$  mean, respectively, either  $a$  or  $a+$ ,  $b$  or  $b-$ . Then if  $\alpha(b-)$  and  $\int_A^{b-} f(x) d\alpha(x)$  exist, and if  $f(x)$  is non-decreasing, then

$$f(b-) \text{ l.b. } [\alpha(b-) - \alpha(x)] \leq \int_A^{b-} f(x) d\alpha(x) \leq f(b-) \text{ u.b. } [\alpha(b-) - \alpha(x)];$$

if  $\alpha(a+)$  and  $\int_{a+}^B f(x) d\alpha(x)$  exist, and if  $f(x)$  is non-increasing, then

$$f(a+) \text{ l.b. } [\alpha(x) - \alpha(a+)] \leq \int_{a+}^B f(x) d\alpha(x) \leq f(a+) \text{ u.b. } [\alpha(x) - \alpha(a+)].$$

Let us consider the case  $b < \infty$ ,  $A = a$ ; the details for the other cases are similar and may be left to the reader. We determine, for  $\epsilon > 0$ , a function  $\delta = \delta(\epsilon) > 0$  such that

$$|\alpha(b-) - \alpha(b-\delta)| = \left| \int_{b-\delta}^{b-} d\alpha(t) \right| \leq \epsilon, \quad \lim_{\epsilon \rightarrow 0} \delta(\epsilon) = 0.$$

Then by use of Lemma 2.1,

$$\begin{aligned}\int_a^{b-\delta} f(t) d\alpha(t) &\geq f(b-\delta) \text{ l.b. } \int_x^{b-\delta} d\alpha(t) \\ &\geq f(b-\delta) \text{ l.b. } \left[ \int_x^{b-} d\alpha(t) - \int_{b-\delta}^{b-} d\alpha(t) \right] \\ &\geq f(b-\delta) \text{ l.b. } \int_x^{b-} d\alpha(t) - \epsilon f(b-\delta).\end{aligned}$$

\* If  $a = -\infty$ ,  $a+\epsilon$  means  $-\epsilon^{-1}$ ; if  $b = +\infty$ ,  $b-\epsilon$  means  $\epsilon^{-1}$ .

Let  $\epsilon \rightarrow 0$ ; then we obtain our inequality in one sense. The opposite inequality is obtained similarly.

LEMMA 2.3. Let  $\phi(t)$  have bounded variation on  $a \leq x \leq b$  for every  $b > a$ ; let  $\phi(\infty)$  exist. Then if  $\int_a^b \psi(t) d\phi(t)$  exists for every  $b > a$ , and if  $\psi(t)$  is, for  $t$  greater than some  $t_0$ , non-negative, monotonic, and bounded, the integral  $\int_a^\infty \psi(t) d\phi(t)$  converges. If  $\psi(t)$  depends on a parameter, and if  $t_0$  and the bound for  $\psi(t)$  are independent of the parameter, the convergence is uniform with respect to the parameter.

In the applications which we shall make, the lemma will usually occur with  $\phi(t)$  an integral. When  $\phi(t)$  is a step-function, the lemma reduces to "Abel's test" for infinite series.\*

The lemma is a simple consequence of Lemma 2.2. Assume  $0 \leq \psi(t) < B, (t > t_0)$ . Given  $\epsilon > 0$ , choose  $T$  so large that for  $T'' > T' \geq T$ ,

$$|\phi(T'') - \phi(T')| < \epsilon B^{-1}.$$

Take  $S' > S > T$ . Then

$$\left| \int_S^{S'} \psi(t) d\phi(t) \right| \leq \begin{cases} \psi(S') \text{ u.b. } |\phi(S') - \phi(S'')|, \\ \psi(S) \text{ u.b. } |\phi(S'') - \phi(S)|, \end{cases}$$

according as  $\psi(t)$  is non-decreasing or non-increasing, respectively; and

$$\left| \int_S^{S'} \psi(t) d\phi(t) \right| < \epsilon, \quad S' > S > T.$$

This establishes the stated convergence.

LEMMA 2.4. Let  $\phi(t^{-1}), \psi(t^{-1})$  satisfy the conditions of Lemma 2.3 with  $a > 0$ . Then the integral  $\int_{0+}^{1/a} \psi(t) d\phi(t)$  exists.

This is reduced to Lemma 2.3 by the change of variable  $t = u^{-1}$ .

One simple application of Lemmas 2.3 and 2.4 is worth stating separately.

LEMMA 2.5. If the  $S_2$  transform (1.3) converges, the integrals

$$(2.1) \quad \int_{0+}^{1/2} \frac{d\alpha(t)}{1-t}, \quad \int_2^\infty \frac{d\alpha(t)}{1-t}$$

converge.

If (1.3) converges, it converges for  $x=1$ . Since the functions  $-1/\log t$  and  $1/\log t$  are positive, monotonic, and bounded on  $(0, 1/2)$  and on  $(2, \infty)$ , re-

\* K. Knopp, *Theory and Application of Infinite Series*, 1928, p. 314.

spectively, the convergence of (2.1) follows by Lemmas 2.4 and 2.3 from the convergence of

$$\int_{0+}^{\infty} \frac{\log t}{1-t} d\alpha(t).$$

**3. The determining function of an  $S_2$  transform.** We prove the following theorem:

**THEOREM 3.1.** *If (1.3) converges, then  $\alpha(0+)$  exists, and*

$$(3.1) \quad \alpha(t) - \alpha(0+) = o(-1/\log t), \quad t \rightarrow 0,$$

$$(3.2) \quad \int_t^{\infty} u^{-1} d\alpha(u) = o(1/\log t), \quad t \rightarrow \infty;$$

$$(3.3) \quad \alpha(t) = o(t/\log t), \quad t \rightarrow \infty.$$

From Theorem 1.1 and Lemma 2.5 we see that

$$\int_{0+}^{1/2} \frac{\log t}{1-t} d\alpha(t), \quad \int_2^{\infty} \frac{\log t}{1-t} d\alpha(t), \quad \int_{0+}^{1/2} \frac{d\alpha(t)}{1-t}, \quad \int_2^{\infty} \frac{d\alpha(t)}{1-t}$$

converge; a simple application of Lemmas 2.3 and 2.4 then shows that

$$(3.4) \quad \int_{0+}^1 d\alpha(t),$$

$$(3.5) \quad \int_{0+}^1 \log t d\alpha(t),$$

$$(3.6) \quad \int_1^{\infty} t^{-1} \log t d\alpha(t),$$

$$(3.7) \quad \int_1^{\infty} t^{-1} d\alpha(t)$$

exist. The existence of (3.4) implies the existence of  $\alpha(0+)$ . If we then write

$$\alpha(t) - \alpha(0+) = \int_{0+}^t d\alpha(u) = \int_{0+}^t \frac{d\beta(u)}{\log u}, \quad \beta(t) = \int_{0+}^t \log u d\alpha(u),$$

we have, because (3.5) converges,

$$\beta(t) = o(1), \quad t \rightarrow 0,$$

$$\begin{aligned} \alpha(t) - \alpha(0+) &= \frac{\beta(t)}{\log t} - \int_{0+}^t \beta(u) d(1/\log u), \quad 0 < t < 1, \\ &= o(-1/\log t), \quad t \rightarrow 0. \end{aligned}$$

This is (3.1).

Since (3.7) converges,

$$\int_t^\infty t^{-1} d\alpha(t) = \int_t^\infty \frac{1}{\log u} d\beta(u), \quad \beta(t) = \int_t^\infty u^{-1} \log u d\alpha(u).$$

The convergence of (3.6) implies that  $\beta(\infty) = 0$ ; hence

$$\begin{aligned} \int_t^\infty t^{-1} d\alpha(t) &= \frac{\beta(t)}{\log t} - \int_t^\infty \beta(u) d(1/\log u), & t > 1, \\ &= o(1/\log t), & t \rightarrow \infty. \end{aligned}$$

This is (3.2), and (3.3) may be obtained from it; or we may proceed as follows.

Because

$$\int_{0+}^\infty \frac{\log t d\alpha(t)}{1+t}$$

converges,

$$\beta(t) = \int_2^t \log u d\alpha(u) = o(t), \quad t \rightarrow \infty.*$$

Then with  $\beta(2) = 0$ ,

$$\begin{aligned} \alpha(t) - \alpha(2) &= \int_2^t \frac{\log u d\alpha(u)}{\log u} = \int_2^t \frac{d\beta(u)}{\log u} \\ &= \frac{\beta(t)}{\log t} - \int_2^t \beta(u) d(1/\log u) = o(t/\log t), \quad t \rightarrow \infty. \end{aligned}$$

**4. Properties of the Stieltjes transform.** The Stieltjes transform in its usual form

$$(4.1) \quad f(x) = \int_0^\infty \frac{d\alpha(t)}{x+t}$$

assumes  $\alpha(t)$  of bounded variation in  $0 \leq t \leq R$  for every positive  $R$ . We shall need to consider also the transform

$$(4.2) \quad f(x) = \int_{0+}^\infty \frac{d\alpha(t)}{x+t} = \lim_{\epsilon \rightarrow 0+, R \rightarrow \infty} \int_\epsilon^R \frac{d\alpha(t)}{x+t},$$

where  $\alpha(t)$  is of bounded variation in  $(\epsilon, R)$  if only  $0 < \epsilon < R < \infty$ . For example, if  $\alpha(t) = t \sin(t^{-1})$  when  $0 < t < 1$ , and if  $\alpha(t) = 0$  when  $t = 0$  and when  $1 \leq t \leq \infty$ , then (4.2) exists although (4.1) is undefined. On the other hand  $f(x) = x^{-1}$  can have the representation (4.1) but not (4.2).

\* D. V. Widder, *The Stieltjes transform*, these Transactions, vol. 43 (1938), pp. 7-60.

By an obvious change of variable we have

$$\int_{0+}^{\infty} \frac{d\alpha(t)}{x+t} = \int_1^{\infty} \frac{d\alpha(t)}{x+t} - \frac{1}{x} \int_1^{\infty} \frac{t d\alpha(1/t)}{t+x^{-1}}.$$

The first integral on the right is in the form (4.1); the second is also except that  $x$  has been replaced by its reciprocal. This enables us to derive easily the facts we need concerning (4.2) from the known results about (4.1).<sup>\*</sup> In particular we showed in §3 that the convergence of (4.2) at  $x=1$  implies the existence of  $\alpha(0+)$ .

We summarize what we shall need in the following theorem:

**THEOREM 4.1.** *If (4.2) converges for some  $x_0 > 0$ , it converges for every  $x > 0$ , and converges uniformly on any interval  $R \geq x \geq \delta$ , ( $0 < \delta < R < \infty$ );  $f(x)$  is analytic for  $x > 0$ , and its derivatives may be evaluated by Leibniz' rule; furthermore,*

$$(4.3) \quad \alpha(0+) \text{ exists;}$$

$$(4.4) \quad \alpha(t) = o(t), \quad t \rightarrow \infty;$$

$$(4.5) \quad f^{(n)}(x) = o(x^{-n-1}), \quad x \rightarrow 0, n = 0, 1, 2, \dots;$$

$$(4.6) \quad f^{(n)}(x) = o(x^{-n}), \quad x \rightarrow \infty, n = 0, 1, 2, \dots.$$

**5. The  $S_2$  transform as an iterated Stieltjes transform.** The  $S_2$  transform was obtained by formally changing the order of integration in

$$(5.1) \quad \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}.$$

In this section we shall show that this formal process is not always permissible; that is, that (5.1) may converge when (1.3) does not. We shall show, however, that when (1.3) converges, (5.1) also converges, and we shall obtain necessary and sufficient conditions for the convergence of (5.1) to imply that of (1.3).

**DEFINITION 5.1.** *Let  $\alpha(t)$  be a normalized function, of bounded variation on every  $(\epsilon, R)$ , ( $0 < \epsilon < R < \infty$ ). Then the iterated integral (5.1), if it exists, is called an iterated Stieltjes transform;  $\alpha(t)$  is called its determining function.*

**LEMMA 5.1.** *The function*

$$H(u) = \frac{1}{\log u} \log \frac{u}{u+\epsilon}$$

*decreases if  $0 < \epsilon < e^{-1}$  and  $0 < u < e^{-1} - \epsilon$ .*

<sup>\*</sup> G. H. Hardy and J. E. Littlewood, *Notes on the theory of series* (XI): On Tauberian theorems, Proceedings of the London Mathematical Society, (2), vol. 30 (1930), pp. 23-37; D. V. Widder, op. cit.

For the proof, we have

$$\begin{aligned} H'(u) &= \frac{1}{\log u} \left( \frac{1}{u} - \frac{1}{u+\epsilon} \right) - \frac{1}{u(\log u)^2} \log \frac{u}{u+\epsilon} \\ &= \frac{-1}{u \log u} \left\{ \frac{u}{u+\epsilon} - \frac{\log(u+\epsilon)}{\log u} \right\}, \end{aligned}$$

which has the sign of the expression in the braces. For  $0 < u < e^{-1}$ , the function  $u \log u^{-1}$  increases; and if  $0 < u + \epsilon < e^{-1}$ , then

$$\begin{aligned} (u + \epsilon) \log \frac{1}{u + \epsilon} &> u \log \frac{1}{u}, \\ \frac{\log(u + \epsilon)}{\log u} &> \frac{u}{u + \epsilon}, \end{aligned}$$

and we have  $H'(u) < 0$ .

THEOREM 5.2. *If the  $S_2$  transform*

$$(5.2) \quad \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t)$$

*converges, then the iterated Stieltjes transform*

$$\int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

*converges, and the two are equal.*

By Lemma 2.5 and an application of Lemmas 2.3 and 2.4, it can be shown that

$$\int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

converges; by Theorem 4.1, it converges uniformly on  $(\epsilon, R)$ ,  $(0 < \epsilon < R < \infty)$ . Hence for  $x > 0$ ,

$$\begin{aligned} \int_{\epsilon}^R \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u} &= \int_{0+}^{\infty} d\alpha(u) \int_{\epsilon}^R \frac{dt}{(x+t)(u+t)} \\ &= \int_{0+}^{\infty} \log \frac{x+R}{u+R} \frac{d\alpha(u)}{u-x} - \int_{0+}^{\infty} \log \frac{x+\epsilon}{u+\epsilon} \frac{d\alpha(u)}{u-x} \\ &= I + J. \end{aligned}$$

We shall show that  $\lim_{R \rightarrow \infty} I = 0$ ,  $\lim_{\epsilon \rightarrow 0} J = f(x)$ .

We take any  $x > 0$  and fix it throughout the discussion. Then

$$\begin{aligned}
 J - f(x) &= \int_{0+}^{\infty} \frac{1}{x-u} \log \left( \frac{x+\epsilon}{x} \frac{u}{u+\epsilon} \right) d\alpha(u) \\
 (5.3) \quad &= \log \frac{x+\epsilon}{x} \int_{0+}^{x-\delta} \frac{d\alpha(u)}{x-u} - \left( \int_{0+}^{\zeta} + \int_{\zeta}^{x-\delta} \right) \log \frac{u+\epsilon}{u} \frac{d\alpha(u)}{x-u} \\
 &\quad + \left( \int_{x-\delta}^{x+\delta} + \int_{x+\delta}^{\infty} \right) \log \left( \frac{x+\epsilon}{x} \frac{u}{u+\epsilon} \right) \frac{d\alpha(u)}{x-u} \\
 &= J_1 + J_2 + J_3 + J_4 + J_5, \quad 0 < \zeta < x - \delta < x.
 \end{aligned}$$

In  $J_4$

$$\begin{aligned}
 \left| \frac{1}{x-u} \log \left( \frac{x+\epsilon}{x} \frac{u}{u+\epsilon} \right) \right| &= \left| \frac{1}{x-u} \int_u^x \left( \frac{1}{t+\epsilon} - \frac{1}{t} \right) dt \right| \\
 &\leq \frac{\epsilon}{(x-\delta)(x-\delta+\epsilon)}.
 \end{aligned}$$

Fix  $\delta$ , ( $0 < \delta < x$ ); then

$$|J_4| \leq \frac{\epsilon}{(x-\delta)(x-\delta+\epsilon)} \int_{x-\delta}^{x+\delta} |d\alpha(u)| = o(1), \quad \epsilon \rightarrow 0.$$

With fixed  $\delta$ , it is clear that  $J_1 = o(1)$ , ( $\epsilon \rightarrow 0$ ).

For  $u > 0$ ,

$$\log \left( \frac{x+\epsilon}{x} \frac{u+\epsilon}{u} \right)$$

is a positive, decreasing function of  $u$ ; since

$$\int_{x+\delta}^{\infty} \frac{d\alpha(u)}{x-u}$$

converges, the integral  $J_5$  converges, by Lemma 2.3, uniformly with respect to  $\epsilon$ , ( $0 < \epsilon < 1$ ). Then we may let  $\epsilon \rightarrow 0$  under the integral sign in  $J_5$  and thus obtain  $J_5 = o(1)$ , ( $\epsilon \rightarrow 0$ ).

Since  $\epsilon \rightarrow 0$ , we may suppose that  $\epsilon < (2e)^{-1}$ . Let  $\zeta = (2e)^{-1}$ . Then  $\zeta + \epsilon < 1/e$ , and by Lemma 5.1,  $H(u)$  decreases for  $0 < u < \zeta$ . But  $H(u) \geq 0$ ,  $H(u)$  is bounded (uniformly with respect to  $\epsilon$  for  $0 < \epsilon < (2e)^{-1}$ ,  $\lim_{\epsilon \rightarrow 0} H(u) = 0$ , and

$$J_2 = \int_{0+}^{\zeta} H(u) \frac{\log(1/u)}{x-u} d\alpha(u).$$

By Lemma 2.4, this integral converges uniformly with respect to  $\epsilon$ , since

$$\int_{0+}^{\xi} \frac{\log(1/u)}{x-u} d\alpha(u)$$

converges, and we may therefore let  $\epsilon \rightarrow 0$  under the integral sign. Hence  $J_2 = o(1)$ , ( $\epsilon \rightarrow 0$ ).

Finally,  $\log[(u+\epsilon)/u]$  is positive decreasing in  $J_3$ ; and

$$\begin{aligned} |J_3| &\leq \log \frac{\xi + \epsilon}{\xi} \quad \text{u.b.} \quad \left| \int_{\xi}^{\xi'} \frac{d\alpha(u)}{x-u} \right| \\ &= o(1), \end{aligned} \quad \epsilon \rightarrow 0.$$

To show that  $I \rightarrow 0$ , we set  $x = y^{-1}$ ,  $R = \eta^{-1}$ , and  $t = u^{-1}$ . Then

$$I = \int_{0+}^{\infty} \log \left( \frac{\eta + y}{\eta + u} \frac{u}{y} \right) \frac{d\beta(u)}{y-u}, \quad \beta(u) = -y \int_{0+}^u t d\alpha(t^{-1}),$$

which is an expression of the same form as (5.3); the  $S_2$  transform

$$\int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\beta(t)$$

converges if (5.2) does, as the change of variable  $u = t^{-1}$  shows. Hence  $\lim_{R \rightarrow \infty} I = 0$ .

**THEOREM 5.3.** *If the iterated Stieltjes transform*

$$(5.4) \quad \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

*converges, the  $S_2$  transform*

$$\int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t)$$

*converges (and is equal to (5.4)) if and only if*

$$(5.5) \quad \alpha(t) - \alpha(0+) = o(-1/\log t), \quad t \rightarrow 0,$$

$$(5.6) \quad \int_t^{\infty} u^{-1} d\alpha(u) = o(1/\log t), \quad t \rightarrow \infty.$$

If the  $S_2$  transform converges, it is equal to (5.4) by the previous theorem; conditions (5.5) and (5.6) are satisfied, by Theorem 3.1.

To establish the converse, we take any fixed  $x > 0$  and consider separately

$$(5.7) \quad \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^x \frac{d\alpha(u)}{t+u},$$

$$(5.8) \quad \int_{0+}^{\infty} \frac{dt}{x+t} \int_x^{\infty} \frac{d\alpha(u)}{t+u}.$$

For each of these integrals, we must show that the order of integration can be changed. We need to do this only for (5.8). For, in (5.7) let us set  $t=s^{-1}$ ,  $u=v^{-1}$ ,  $x=y^{-1}$ . Then (5.7) becomes

$$(5.9) \quad -y \int_{0+}^{\infty} \frac{ds}{y+s} \int_v^{\infty} \frac{v d\alpha(v^{-1})}{s+v},$$

which has the same form as (5.8), if we write

$$\beta(u) = -y \int_v^u v d\alpha(v^{-1}), \quad \beta(y) = 0;$$

also

$$\int_t^{\infty} u^{-1} d\beta(u) = y\alpha(t^{-1}) = o(1/\log t), \quad t \rightarrow \infty,$$

if (5.5) is satisfied. Thus if we may change the order of integration in (5.8) when (5.6) is satisfied, then we may change the order in (5.9) if (5.5) is satisfied, and hence in (5.7).

We consider

$$I(R) = \int_{0+}^R \frac{dt}{x+t} \int_x^{\infty} \frac{d\alpha(u)}{t+u},$$

which by hypothesis approaches a limit as  $R \rightarrow \infty$ . The integral

$$\int_x^{\infty} \frac{d\alpha(u)}{t+u}$$

converges uniformly for  $0 \leq t \leq R$ , as one sees by applying Theorem 4.1, after setting  $u=v+x$ ,  $t=s-x$ . Therefore

$$\begin{aligned} I(R) &= \int_x^{\infty} d\alpha(u) \int_0^R \frac{dt}{(x+t)(u+t)} \\ &= \int_x^{\infty} \left\{ \log \frac{x+R}{u+R} + \log \frac{u}{x} \right\} \frac{d\alpha(u)}{u-x}. \end{aligned}$$

We wish to show that

$$J(R) = \int_x^R \frac{\log(u/x)}{u-x} d\alpha(u)$$

has the same limit when  $R \rightarrow \infty$  as  $I(R)$ . We consider, for  $R > 2x$ , the difference

$$\begin{aligned}
 I(R) - J(R) &= \log \frac{x+R}{x} \int_R^\infty \frac{d\alpha(u)}{u-x} - \int_R^\infty \log \frac{u+R}{u} \frac{d\alpha(u)}{u-x} \\
 &\quad + \left( \int_x^{2x} + \int_{2x}^R \right) \log \frac{x+R}{u+R} \frac{d\alpha(u)}{u-x} \\
 &= I_1 + I_2 + I_3 + I_4.
 \end{aligned}$$

We note that the integral  $I_1$  converges by (5.6) combined with Lemma 2.3. Using Lemma 2.2 and applying (5.6), we obtain

$$\begin{aligned}
 |I_1| &= O(\log R) \left| \int_R^\infty \frac{u}{u-x} \frac{d\alpha(u)}{u} \right| \\
 &\leq O(\log R) \frac{R}{R-x} \text{u.b.}_{R' \geq R} \left| \int_R^{R'} \frac{d\alpha(u)}{u} \right| \\
 &= o(1), \quad R \rightarrow \infty.
 \end{aligned}$$

It is a simple consequence of (5.6) that

$$(5.10) \quad \beta(t) \equiv \int_t^\infty \frac{d\alpha(u)}{u-x} = o(1/\log t), \quad t \rightarrow \infty.$$

Therefore, since  $\log [(u+R)/u]$  is positive decreasing, ( $u > R$ ), we have

$$|I_2| \leq \log 2 \text{u.b.}_{R' \geq R} \left| \int_R^{R'} \frac{d\alpha(u)}{u-x} \right| = o(1), \quad R \rightarrow \infty.$$

Also,

$$\begin{aligned}
 I_4 &= \int_{2x}^R \log \frac{x+R}{u+R} d\beta(u) \\
 &= \beta(R) \log \frac{x+R}{2R} - \beta(2x) \log \frac{x+R}{2x+R} + \int_{2x}^R \frac{\beta(u)}{u+R} du \\
 &= o(1), \quad R \rightarrow \infty,
 \end{aligned}$$

by use of (5.10).

Finally,

$$\begin{aligned}
 |I_3| &= \left| \int_x^{2x} \frac{\log(x+R) - \log(u+R)}{x-u} d\alpha(u) \right| \\
 &\leq \frac{1}{x+R} \int_x^{2x} |d\alpha(u)| = o(1), \quad R \rightarrow \infty.
 \end{aligned}$$

To complete our discussion of the relations between the  $S_2$  transform and

the iterated Stieltjes transform, we need to establish the following theorem:

**THEOREM 5.4.** *There exists a function  $\alpha(t)$  of bounded variation on  $(0, \infty)$ , such that*

$$(5.11) \quad \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

converges, and

$$(5.12) \quad \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t)$$

diverges.

Let  $\{u_n\}$ ,  $\{u'_n\}$ ,  $(n=1, 2, \dots)$ , be sequences of points such that  $0 < u'_{n+1} < u_n < u'_n < 1$ , and such that the series

$$\sum_{n=1}^{\infty} \frac{1}{\log u_n}, \quad \sum_{n=1}^{\infty} \frac{\log u'_n - \log u_n}{\log u_n}$$

converge (for example,  $u_n = 2^{-n^3}$ ,  $u'_n = 2^{-n^3+n}$ ). Then the function

$$\alpha(u) = \begin{cases} -(\log u_n)^{-1}, & u_n < u < u'_n, \\ -(2 \log u_n)^{-1}, & u = u_n, \quad u = u'_n, \\ 0, & \text{elsewhere,} \end{cases}$$

has the desired properties.

The total variation of  $\alpha(u)$  on  $(0, \infty)$  is

$$\sum_{n=1}^{\infty} \frac{-2}{\log u_n} < \infty.$$

Moreover,

$$\int_{0+}^1 \frac{\alpha(u)}{u} du = \sum_{n=1}^{\infty} \frac{-1}{\log u_n} \int_{u_n}^{u'_n} \frac{du}{u} = \sum_{n=1}^{\infty} \frac{\log u_n - \log u'_n}{\log u_n} < \infty.$$

Also,  $\alpha(u) \geq 0$ ;  $\alpha(0) = \alpha(0+) = \alpha(1) = 0$ ;  $\alpha(u) - \alpha(0+) \neq o(-1/\log u)$ ,  $(u \rightarrow 0)$ , since  $\alpha(u_n) = (-2 \log u_n)^{-1}$  (but  $\alpha(u) = O(-1/\log u)$ ,  $(u \rightarrow 0)$ ), and

$$\begin{aligned} \int_{0+}^1 \frac{\alpha(u)}{u} du &= \int_{0+}^1 \alpha(u) du \int_0^{\infty} \frac{dt}{(u+t)^2} \\ &= \int_0^{\infty} dt \int_{0+}^1 \frac{\alpha(u) du}{(u+t)^2} \\ &= \int_0^{\infty} dt \int_{0+}^1 \frac{d\alpha(u)}{u+t} \end{aligned}$$

(the change of order of integration is legitimate because the integrand is non-negative). Hence, by use of Lemmas 2.3 and 2.4, (5.11) converges. But (5.12) must diverge, since (3.1) is not satisfied.

## CHAPTER II. INVERSION OF THE TRANSFORMS

**6. The inversion operator.** In the remainder of this paper, unless the contrary is specified, all quantities are to be real, and the domain of all functions is  $(0, \infty)$ .

**DEFINITION 6.1.\*** For a function  $f(x)$  of class  $C^{2k-1}$ , an operator  $L_{k,z}[f(x)]$  is defined by

$$(6.1) \quad L_{k,z}[f(x)] = c_k(-x)^{k-1}[x^k f(x)]^{(2k-1)}, \quad k = 1, 2, \dots,$$

$$(6.2) \quad c_1 = 1, \quad c_k = \frac{1}{k!(k-2)!}, \quad k = 2, 3, \dots$$

**DEFINITION 6.2.** For a function  $f(x)$  of class  $C^{4k-2}$ , an operator  $H_{k,z}[f(x)]$  is defined by

$$(6.3) \quad H_{k,z}[f(x)] = L_{k,z}\{L_{k,z}[f(x)]\}.$$

**THEOREM 6.1.** If  $f(x)$  is an iterated Stieltjes transform

$$(6.4) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u},$$

then

$$(6.5) \quad H_{k,z}[f(x)] = \int_{0+}^{\infty} F_k(u, x) d\alpha(u), \quad k = 2, 3, \dots,$$

where

$$(6.6) \quad F_k(u, x) = d_k^2 x^{k-1} u^k \int_0^{\infty} \frac{t^{2k-1} dt}{(x+t)^{2k}(u+t)^{2k}},$$

$$(6.7) \quad d_k = (2k-1)! c_k.$$

We have

$$f(x) = \int_{0+}^{\infty} \frac{\phi(t) dt}{x+t}, \quad \phi(t) = \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u};$$

by Theorem 4.1,  $f(x)$  and  $\phi(t)$  are of class  $C^\infty$ ; their derivatives may be evaluated by Leibniz' rule; and, for  $n=0, 1, 2, \dots$ ,

$$(6.8) \quad \begin{aligned} \phi^{(n)}(t) &= o(t^{-n-1}), & t \rightarrow 0, \\ \phi^{(n)}(t) &= o(t^{-n}), & t \rightarrow \infty. \end{aligned}$$

\* D. V. Widder, op. cit.

Now for  $k \geq 2$

$$\begin{aligned} \frac{1}{k!} [x^k f(x)]^{(k)} &= \frac{1}{k!} \int_{0+}^{\infty} \frac{\partial^k}{\partial x^k} \left( \frac{x^k}{x+t} \right) \phi(t) dt \\ &= \int_0^{\infty} \frac{t^k \phi(t)}{(x+t)^{k+1}} dt. \end{aligned}$$

If we integrate by parts  $k$  times, we find, by (6.8), that the integrated terms all vanish, so that

$$[x^k f(x)]^{(k)} = \int_0^{\infty} \frac{[t^k \phi(t)]^{(k)}}{x+t} dt.$$

Thus

$$\begin{aligned} x^{k-1} [x^k f(x)]^{(2k-1)} &= (-1)^{k-1} (k-1)! \int_0^{\infty} \frac{x^{k-1} [t^k \phi(t)]^{(k)}}{(x+t)^k} dt \\ &= (-1)^{k-1} \int_0^{\infty} [t^k \phi(t)]^{(k)} \frac{\partial^{k-1}}{\partial t^{k-1}} \left( \frac{t^{k-1}}{x+t} \right) dt. \end{aligned}$$

We integrate by parts  $k-1$  times; the integrated terms all vanish, and we obtain

$$(6.9) \quad L_{k,z}[f(x)] = \int_0^{\infty} \frac{L_{k,t}[\phi(t)]}{x+t} dt.$$

If

$$g(x) = \int_{0+}^{\infty} \frac{d\psi(t)}{x+t},$$

then

$$L_{k,z}[g(x)] = d_k x^{k-1} \int_{0+}^{\infty} \frac{t^k d\psi(t)}{(x+t)^{2k}}.$$

Applying this formula twice to (6.9), we obtain

$$\begin{aligned} (6.10) \quad H_{k,z}[f(x)] &= d_k x^{k-1} \int_0^{\infty} \frac{t^k L_{k,t}[\phi(t)] dt}{(x+t)^{2k}} \\ &= d_k^2 x^{k-1} \int_0^{\infty} \frac{t^{2k-1} dt}{(x+t)^{2k}} \int_{0+}^{\infty} \frac{u^k d\alpha(u)}{(u+t)^{2k}}. \end{aligned}$$

We now wish to change the order of integration in (6.10). Writing

$$A(u, t) = \frac{u^k}{(u+t)^{2k}}$$

we see, by use of (4.3) and (4.4), that

$$\int_{0+}^{\infty} [\alpha(u) - \alpha(0+)] \frac{\partial}{\partial u} A(u, t) du$$

exists, and that

$$(6.11) \quad - \int_{0+}^{\infty} [\alpha(u) - \alpha(0+)] \frac{\partial}{\partial u} A(u, t) du = \int_{0+}^{\infty} A(u, t) d[\alpha(u) - \alpha(0+)] \\ = \int_{0+}^{\infty} \frac{u^k d\alpha(u)}{(t+u)^{2k}}.$$

Now consider the integral

$$d_k^2 x^{k-1} \int_{0+}^{\infty} u^k d\alpha(u) \int_0^{\infty} \frac{t^{2k-1} dt}{(x+t)^{2k}(u+t)^{2k}} = \int_{0+}^{\infty} F_k(u, x) d[\alpha(u) - \alpha(0+)],$$

$k \geq 2$ . For we anticipate the inequalities of Lemma 7.3; combined with (4.3) and (4.4), they show that the last integral exists and is equal to

$$- \int_{0+}^{\infty} [\alpha(u) - \alpha(0+)] \frac{\partial}{\partial u} F_k(u, x) du \\ = - k d_k^2 x^{k-1} \int_{0+}^{\infty} [\alpha(u) - \alpha(0+)] u^{k-1} du \int_0^{\infty} \frac{t^{2k-1}(t-u) dt}{(x+t)^{2k}(u+t)^{2k+1}}.$$

The repeated integral is easily seen (compare (7.5), (7.6)) to be dominated by

$$k \int_0^{\infty} |\alpha(u) - \alpha(0+)| u^{-1} F_k(u, x) du,$$

which (because of (4.3), (4.4), and Lemma 7.3) converges ( $k \geq 2$ ); hence the order of integration in the repeated integral can be changed, so that

$$\int_{0+}^{\infty} F_k(u, x) d\alpha(u) = - d_k^2 x^{k-1} \int_0^{\infty} \frac{t^{2k-1} dt}{(x+t)^{2k}} \int_{0+}^{\infty} [\alpha(u) - \alpha(0+)] \frac{\partial}{\partial u} A(u, t) du.$$

Referring to (6.11), we obtain (6.5).

COROLLARY 6.1.1. For  $k \geq 2$ ,

$$H_{k,x} \left[ \frac{\log(x/t)}{x-t} \right] = F_k(t, x).$$

We may write

$$\frac{\log(x/t)}{x-t} = \int_{0+}^{\infty} \frac{\log(x/u)}{x-u} d\alpha(u), \quad \alpha(u) = \begin{cases} 0, & 0 \leq u < t, \\ 1/2, & u = t, \\ 1, & u > t. \end{cases}$$

Then by Theorem 5.2

$$\frac{\log(x/t)}{x-t} = \int_{0+}^{\infty} \frac{du}{x+u} \int_{0+}^{\infty} \frac{d\alpha(v)}{t+v};$$

by Theorem 6.1

$$H_{k,z} \left[ \frac{\log(x/t)}{x-t} \right] = \int_{0+}^{\infty} F_k(u, x) d\alpha(u) = F_k(t, x), \quad k \geq 2.$$

**7. Properties of the function  $F_k(u, x)$ .** We have the following lemma:

**LEMMA 7.1.** *If  $m=1, 2, \dots; n=2, 3, \dots; n>m$ , then*

$$\int_0^{\infty} \frac{u^{m-1} du}{(t+u)^n} = \frac{(m-1)!(n-m-1)!}{t^{n-m}(n-1)!}.$$

This is the familiar formula for the beta function.

**LEMMA 7.2.** *If  $k=2, 3, \dots$ , then*

$$\int_0^{\infty} F_k(u, x) du = 1, \\ x \int_0^{\infty} u^{-1} F_k(u, x) du = \int_0^{\infty} F_k(u, x) dx = \left( \frac{k-1}{k} \right)^2.$$

These formulas are obtained by applying Lemma 7.1 twice to each of the repeated integrals in question, after changing the order of integration.

**LEMMA 7.3.** *If  $k=2, 3, \dots$  and  $x>0$  is fixed, then*

$$(7.1) \quad F_k(u, x) = O(u^{-k+1}), \quad u \rightarrow \infty,$$

$$(7.2) \quad F_k(u, x) = O(u^{k-1}), \quad u \rightarrow 0,$$

$$(7.3) \quad \frac{\partial}{\partial u} F_k(u, x) = O(u^{-k}), \quad u \rightarrow \infty,$$

$$(7.4) \quad \frac{\partial}{\partial u} F_k(u, x) = O(u^{k-2}), \quad u \rightarrow 0.$$

Since  $u/(u+t) < 1$ ,  $1/(u+t) < t^{-1}$ , and  $t/(u+t) < 1$ , ( $u>0, t>0$ ), we have

$$u^{k-1} F_k(u, x) \leq d_k^2 x^{k-1} \int_0^{\infty} \frac{t^{2k-2} dt}{(x+t)^{2k}},$$

which gives (7.1). We have also the following relation from which (7.2) follows:

$$F_k(u, x) \leq d_k^2 x^{k-1} u^k \int_0^{\infty} \frac{dt}{(x+t)^{2k}(u+t)} \leq d_k^2 x^{k-1} u^{k-1} \int_0^{\infty} \frac{dt}{(x+t)^{2k}}.$$

Since

$$(7.5) \quad \frac{\partial}{\partial u} F_k(u, x) = k d_k^2 u^{k-1} x^{k-1} \int_0^\infty \frac{t^{2k-1}(t-u)dt}{(x+t)^{2k}(u+t)^{2k+1}},$$

and since for  $u > 0$  and  $t > 0$  the inequality  $|t-u|/(u+t) < 1$  holds, it follows that

$$(7.6) \quad \left| \frac{\partial}{\partial u} F_k(u, x) \right| \leq \frac{k}{u} F_k(u, x);$$

hence (7.3) and (7.4) follow from (7.6) combined with (7.1) and (7.2).

LEMMA 7.4.  $F_k(u, x)$ , as a function of  $u$ , increases for  $u < x$  and decreases for  $u > x$ .

It is sufficient to establish this for  $x=1$ , since  $F_k(u, x)$  is homogeneous of degree  $-1$ , so that

$$F_k(u, x) = x^{-1} F_k(x^{-1}u, 1).$$

We have, from (7.5),

$$\begin{aligned} \frac{1}{k d_k^2} \frac{\partial}{\partial u} F_k(u, 1) &= u^{k-1} \int_0^\infty \frac{t^{2k} dt}{(1+t)^{2k}(u+t)^{2k+1}} - u^k \int_0^\infty \frac{t^{2k-1} dt}{(1+t)^{2k}(u+t)^{2k+1}} \\ &= I_1 - I_2. \end{aligned}$$

If we make the change of variable  $t=u/s$ , and replace  $s$  by  $t$  in the result, we find that

$$\begin{aligned} I_1 &= u^{k-1} \int_0^\infty \frac{t^{2k-1} dt}{(1+t)^{2k+1}(u+t)^{2k}}, \\ I_1 - I_2 &= u^{k-1} \int_0^\infty \frac{t^{2k-1}}{(1+t)^{2k}(u+t)^{2k}} \left[ \frac{1}{1+t} - \frac{u}{u+t} \right] dt \\ &= (1-u) u^{k-1} \int_0^\infty \frac{t^{2k} dt}{(1+t)^{2k+1}(u+t)^{2k+1}}, \end{aligned}$$

which has the sign of  $(1-u)$ .

8. **Some preliminary limits.** We establish the following lemma:

LEMMA 8.1. If  $0 < y < 1$ , then

$$(8.1) \quad \lim_{k \rightarrow \infty} k d_k \int_0^y \frac{u^{k-1} du}{(u+1)^{2k}} = 0.$$

If  $0 < y < 1$ , then the function  $u(u+1)^{-2}$ , which has a single maximum (at  $u=1$ ), increases on  $(0, y)$ , and

$$0 \leq kd_k \int_0^y \frac{u^{k-1} du}{(u+1)^{2k}} < kd_k \frac{y^{k-1}}{(y+1)^{2k-2}} \int_0^y \frac{du}{(u+1)^2};$$

the last expression approaches zero ( $k \rightarrow \infty$ ), since it is the general term of a convergent infinite series. In fact, the test ratio for the series is

$$\frac{2(2k+1)}{k-1} \frac{y}{(y+1)^2}$$

which approaches a limit less than unity ( $k \rightarrow \infty$ ).

LEMMA 8.2. *Let*

$$\begin{aligned} H_k(y) &= \int_0^y F_k(1, x) dx \\ (8.2) \quad &= d_k^2 \int_0^y x^{k-1} dx \int_0^\infty \frac{t^{2k-1} dt}{(x+t)^{2k}(1+t)^{2k}}. \end{aligned}$$

Then

$$(8.3) \quad \lim_{k \rightarrow \infty} k H_k(y) = 0, \quad 0 \leq y < 1,$$

$$(8.4) \quad \lim_{k \rightarrow \infty} k [1 - H_k(y)] = 0, \quad y > 1,$$

$$(8.5) \quad \lim_{k \rightarrow \infty} H_k(1) = \frac{1}{2}.$$

We consider first  $0 < y < 1$ . In (8.2) change the order of integration and make the change of variable  $x = ut$ . Then

$$(8.6) \quad d_k^{-2} H_k(y) = \int_0^\infty dt \int_0^{y/t} \frac{(ut)^{k-1} du}{(u+1)^{2k}(t+1)^{2k}}$$

$$(8.7) \quad \leq \left( \int_0^\infty \int_0^\infty - \int_{y/2}^\infty \int_{y/2}^\infty \right) \frac{u^{k-1} t^{k-1} du dt}{(u+1)^{2k}(t+1)^{2k}},$$

since the integrand is nonnegative, and the domain of integration in (8.7) includes that in (8.6). Thus

$$\begin{aligned} d_k^{-2} H_k(y) &\leq \left\{ \int_0^\infty \frac{t^{k-1} dt}{(t+1)^{2k}} - \int_{y/2}^\infty \frac{t^{k-1} dt}{(t+1)^{2k}} \right\} \\ &\quad \cdot \left\{ \int_0^\infty \frac{t^{k-1} dt}{(t+1)^{2k}} + \int_{y/2}^\infty \frac{t^{k-1} dt}{(t+1)^{2k}} \right\} \\ &\leq 2 \int_0^{y/2} \frac{t^{k-1} dt}{(t+1)^{2k}} \int_0^\infty \frac{u^{k-1} du}{(u+1)^{2k}}; \end{aligned}$$

and

$$H_k(y) \leq 2d_k \int_0^{y^{1/2}} \frac{u^{k-1} du}{(t+1)^{2k}},$$

since by Lemma 7.1 it is seen that

$$d_k \int_0^\infty \frac{u^{k-1} du}{(u+1)^{2k}} = \frac{k-1}{k} < 1.$$

Then by Lemma 8.1,  $\lim_{k \rightarrow \infty} kH_k(y) = 0$ , ( $0 < y < 1$ ).

Now consider  $y > 1$ . We have by Lemma 7.2

$$\begin{aligned} \int_0^\infty F_k(1, x) dx &= \left( \frac{k-1}{k} \right)^2, \\ \left( \frac{k-1}{k} \right)^2 - H_k(y) &= \int_y^\infty F_k(1, x) dx = \int_0^{y^{-1}} x^{-2} F_k(1, x^{-1}) dx. \end{aligned}$$

But

$$(8.8) \quad F_k(1, x) = x^{-1} F_k(x^{-1}, 1) = x^{-2} F_k(1, x^{-1}),$$

and (8.4) thus follows by what has already been established.

Finally consider  $H_k(1)$ . Using (8.8), we obtain

$$\begin{aligned} H_k(1) &= \int_0^1 F_k(1, x) dx = \int_1^\infty x^{-2} F_k(1, x^{-1}) dx = \int_1^\infty F_k(1, x) dx; \\ 2H_k(1) &= \int_0^\infty F_k(1, x) dx = \left( \frac{k-1}{k} \right)^2 \rightarrow 1, \quad k \rightarrow \infty. \end{aligned}$$

LEMMA 8.3. If  $u \neq x$ , then

$$(8.9) \quad \lim_{k \rightarrow \infty} F_k(u, x) = 0.$$

It is sufficient to consider  $F_k(u, 1)$ , because  $F_k(u, x) = x^{-1} F_k(u/x, 1)$ . Since  $F_k(0, 1) = 0$ , we have by (7.6)

$$0 \leq F_k(y, 1) = \int_0^y \frac{\partial}{\partial u} F_k(u, 1) du \leq k \int_0^y u^{-1} F_k(u, 1) du.$$

That is,

$$0 \leq F_k(y, 1) \leq k \int_0^y F_k(1, u) du = kH_k(y) = o(1), \quad k \rightarrow \infty,$$

if  $0 < y < 1$ , by Lemma 8.2. Since

$$F_k(x, u) = xu^{-1} F_k(u, x) = u^{-1} F_k(xu^{-1}, 1),$$

we have  $F_k(y^{-1}, 1) = F_k(y, 1)$ ; and (8.9) for  $u/x < 1$  implies (8.9) for  $u/x > 1$ .

It is interesting to compare Lemma 8.3 with the following lemma:

LEMMA 8.4. *If  $x > 0$ , then*

$$F_k(x, x) \sim \frac{1}{x} \left( \frac{k}{8\pi} \right)^{1/2}, \quad k \rightarrow \infty.$$

For,

$$F_k(x, x) = d_k^2 x^{2k-1} \int_0^\infty \frac{t^{2k-1} dt}{(x+t)^{4k}} = \frac{1}{x} d_k^2 \frac{(2k-1)!(2k-1)!}{(4k-1)!},$$

and an application of Stirling's formula gives the result.

9. A singular integral. We prove the following lemma:

LEMMA 9.1. *If  $u^m \beta(u)$  is bounded and integrable on  $0 \leq u \leq x$  for some integer  $m \geq 0$ , then for  $0 < \delta < x$*

$$I_k = \int_0^{x-\delta} \beta(u) \frac{\partial}{\partial u} F_k(u, x) du \rightarrow 0, \quad k \rightarrow \infty.$$

That the integrand is integrable for sufficiently large  $k$  follows from Lemma 7.3. By (7.6)

$$|I_k| \leq k \int_0^{x-\delta} u^{-1} |\beta(u)| F_k(u, x) du.$$

Set  $u = xv$ , and assume  $|u^m \beta(u)| < B$ , ( $0 \leq u \leq x$ ). For  $k > m+1$ , we have

$$\begin{aligned} |I_k| &\leq k \int_0^{1-x^{-1}\delta} v^{-1} |\beta(xv)| F_k(vx, x) dv \\ &= kx^{-1} \int_0^{1-x^{-1}\delta} v^{-1} |\beta(xv)| F_k(v, 1) dv \\ &\leq x^{-m-1} Bk \int_0^{1-x^{-1}\delta} v^{-m-1} F_k(v, 1) dv \\ &= x^{-m-1} Bk d_k^2 \int_0^{1-x^{-1}\delta} v^{k-m-1} dv \int_0^\infty \frac{t^{2k-1} dt}{(1+t)^{2k}(t+v)^{2k}} \\ &\leq x^{-m-1} Bk d_k^2 \int_0^{1-x^{-1}\delta} v^{k-m-1} dv \int_0^\infty \frac{t^{2k-2m-1} dt}{(1+t)^{2k-2m}(t+v)^{2k-2m}} \\ &= x^{-m-1} Bk d_k^2 d_{k-m}^2 H_{k-m}(1-x^{-1}\delta) \\ &= o(1), \quad k \rightarrow \infty, \end{aligned}$$

by Lemma 8.2, since  $d_k/d_{k-m} = O(1)$ , ( $k \rightarrow \infty$ ).

LEMMA 9.2. If  $\beta(u)$  is bounded and integrable on  $(x, R)$  for every  $R > x > 0$ , and if  $\beta(u) = O(u^m)$ ,  $(u \rightarrow \infty)$ , for some integer  $m \geq 0$ , then for every  $\delta > 0$ ,

$$J_k = \int_{x+\delta}^{\infty} \beta(u) \frac{\partial}{\partial u} F_k(u, x) du \rightarrow 0, \quad k \rightarrow \infty.$$

We proceed as in Lemma 9.1. Assume  $|\beta(u)u^{-m}| < B$  for  $u \geq x$ . For  $k > m+1$ , we then have

$$\begin{aligned} |J_k| &\leq k \int_{x+\delta}^{\infty} u^{-1} |\beta(u)| F_k(u, x) du \\ &\leq kBx^{m-1} \int_{1+x^{-1}\delta}^{\infty} v^{m-1} F_k(v, 1) dv \\ &= kBx^{m-1} d_k^2 \int_{1+x^{-1}\delta}^{\infty} v^{k+m-1} dv \int_0^{\infty} \frac{t^{2k-1} dt}{(1+t)^{2k}(t+v)^{2k}} \\ &\leq kBx^{m-1} d_k^2 \int_{1+x^{-1}\delta}^{\infty} v^{k+m-1} dv \int_0^{\infty} \frac{t^{2k-2m-1} dt}{(1+t)^{2k-2m}(t+v)^{2k-2m}} \\ &= kBx^{m-1} d_k^2 d_{k-m}^{-2} \left\{ \left( \frac{k-m-1}{k-m} \right)^2 - H_{k-m}(1+x^{-1}\delta) \right\} \\ &= o(1), \quad k \rightarrow \infty. \end{aligned}$$

THEOREM 9.3. If  $\phi(u)$  is integrable on every  $(\epsilon, R)$ ,  $(0 < \epsilon < R < \infty)$ , and if there exist integers  $m \geq 0$ ,  $n \geq 0$ , such that

$$(9.1) \quad \gamma(u) \equiv \int_1^u \phi(t) dt = \begin{cases} O(u^{-m}), & u \rightarrow 0, \\ O(u^n), & u \rightarrow \infty; \end{cases}$$

then for  $k$  sufficiently large, the integral

$$(9.2) \quad G_k(x) = \int_{0+}^{\infty} F_k(u, x) \phi(u) du$$

exists, and

$$(9.3) \quad \lim_{k \rightarrow \infty} G_k(x) = \phi(x)$$

for every  $x > 0$  for which either

$$(9.4) \quad \int_x^t |\phi(u) - \phi(x)| du = o(|t-x|), \quad t \rightarrow x,$$

or  $\phi(x+)$  and  $\phi(x-)$  exist with

$$(9.5) \quad \phi(x) = \frac{1}{2}[\phi(x+) + \phi(x-)].$$

We note that (9.4) is satisfied for almost all  $x$ , and, in particular, wherever  $\phi(x)$  is continuous.\*

We show first that (9.2) exists. For  $R > 1$ ,

$$\begin{aligned}\int_1^R F_k(u, x)\phi(u)du &= \int_1^R F_k(u, x)d\gamma(u) \\ &= F_k(R, x)\gamma(R) - \int_1^R \gamma(u) \frac{\partial}{\partial u} F_k(u, x)du.\end{aligned}$$

By Lemma 7.3 and (9.1), if  $k \geq n+2$ , this expression is

$$O(R^{-n-1})O(R^n) - \int_1^R O(u^n)O(u^{-n-2})du,$$

which approaches a limit as  $R \rightarrow \infty$ . A similar argument shows that

$$\int_{0+}^1 F_k(u, x)\phi(u)du, \quad k \geq m+2,$$

converges.

Since by Lemma 7.2  $\int_0^\infty F_k(u, x)du = 1$ , ( $k \geq 2$ ), we have

$$(9.6) \quad D_k(x) \equiv G_k(x) - \phi(x) = \int_{0+}^\infty [\phi(u) - \phi(x)]F_k(u, x)du.$$

Let  $x$  be a point where (9.4) is satisfied, and set

$$\beta(t, x) = \int_x^t [\phi(u) - \phi(x)]du.$$

Then

$$D_k(x) = \int_{0+}^\infty F_k(u, x)d_u\beta(u, x) = - \int_0^\infty \beta(u, x) \frac{\partial}{\partial u} F_k(u, x)du,$$

the integrated terms vanishing (for  $k$  sufficiently large) by (9.1) and Lemma 7.3. Assuming (9.4), we find that  $\beta(u, x) = o(|u-x|)$ , ( $u \rightarrow x$ ); we can therefore choose  $\delta$ , ( $0 < \delta < x$ ), so that,  $\epsilon > 0$  being given,

$$(9.7) \quad |\beta(u, x)| \leq \epsilon |x - u|, \quad |x - u| \leq \delta;$$

then

$$\begin{aligned}-D_k(x) &= \left( \int_0^{x-\delta} + \int_{x-\delta}^{x+\delta} + \int_{x+\delta}^\infty \right) \beta(u, x) \frac{\partial}{\partial u} F_k(u, x)du \\ &= I_1 + I_2 + I_3.\end{aligned}$$

\* See, for example, E. C. Titchmarsh, *The Theory of Functions*, 1932, p. 364.

By using (9.1), we see that  $\beta(u, x)$ , as a function of  $u$ , satisfies the hypotheses of Lemmas 9.1 and 9.2, and hence that

$$\lim_{k \rightarrow \infty} (I_1 + I_3) = 0.$$

Since, by Lemma 7.4,  $F_k(u, x)$  increases for  $u < x$  and decreases for  $u > x$ ,

$$|I_2| \leq \int_{x-\delta}^x |\beta(u, x)| d_u F_k(u, x) - \int_x^{x+\delta} |\beta(u, x)| d_u F_k(u, x);$$

and if we use (9.7),

$$\begin{aligned} |I_2| &\leq \epsilon \int_{x-\delta}^x (x-u) d_u F_k(u, x) - \epsilon \int_x^{x+\delta} (u-x) d_u F_k(u, x) \\ &= -\epsilon \delta F_k(x-\delta, x) - \epsilon \delta F_k(x+\delta, x) + \epsilon \int_{x-\delta}^{x+\delta} F_k(u, x) du \\ &< \epsilon \int_0^\infty F_k(u, x) du = \epsilon, \end{aligned}$$

since  $F_k(u, x) \geq 0$ . Therefore

$$\limsup_{k \rightarrow \infty} |D_k(x)| \leq \epsilon;$$

and since  $\epsilon$  was arbitrary,

$$\lim_{k \rightarrow \infty} G_k(x) = \phi(x).$$

Now suppose that (9.5) is satisfied. Set

$$\begin{aligned} \theta(u) &= \begin{cases} \phi(x-), & u < x, \\ \phi(x), & u = x, \\ \phi(x+), & u > x, \end{cases} \\ \omega(u) &= \phi(u) - \theta(u); \end{aligned}$$

then  $\omega(u)$  is continuous at  $u=x$ , and  $\omega(x)=0$ . Hence

$$G_k(x) = \int_{0+}^\infty F_k(u, x) \omega(u) du + \int_0^\infty F_k(u, x) \theta(u) du;$$

$\omega(u)$  satisfies the hypotheses of the theorem and satisfies (9.4) at  $u=x$ . By what has already been established,

$$\lim_{k \rightarrow \infty} \int_{0+}^\infty F_k(u, x) \omega(u) du = \omega(x) = 0.$$

On the other hand,

$$\begin{aligned}\int_0^\infty F_k(u, x)\theta(u)du &= \phi(x-) \int_0^x F_k(u, x)du + \phi(x+) \int_x^\infty F_k(u, x)du \\ &= x\phi(x-) \int_0^1 F_k(ux, x)du + x\phi(x+) \int_1^\infty F_k(ux, x)du \\ &= \phi(x-) \int_0^1 F_k(u, 1)du + \phi(x+) \int_1^\infty F_k(u, 1)du;\end{aligned}$$

and

$$(9.8) \quad \int_0^\infty F_k(u, x)\theta(u)du = \phi(x+) + [\phi(x-) - \phi(x+)] \int_0^1 F_k(u, 1)du.$$

But

$$\begin{aligned}\int_0^1 F_k(u, 1)du &= \int_0^1 uF_k(1, u)du \\ &= H_k(1) - \int_0^1 (1-u)F_k(1, u)du;\end{aligned}$$

and for  $0 < \epsilon < 1$ ,

$$\begin{aligned}\int_0^1 (1-u)F_k(1, u)du &= \int_0^{1-\epsilon} (1-u)F_k(1, u)du + \int_{1-\epsilon}^1 (1-u)F_k(1, u)du \\ &\leq \int_0^{1-\epsilon} F_k(1, u)du + \epsilon \int_0^\infty F_k(1, u)du \\ &= H_k(1 - \epsilon) + \epsilon \left( \frac{k-1}{k} \right)^2.\end{aligned}$$

By use of Lemma 8.2, we see that

$$\limsup_{k \rightarrow \infty} \int_0^1 (1-u)F_k(1, u)du \leq \epsilon,$$

and hence that

$$\lim_{k \rightarrow \infty} \int_0^1 F_k(u, 1)du = \lim_{k \rightarrow \infty} H_k(1) = 1/2.$$

Thus (9.8) gives

$$\lim_{k \rightarrow \infty} \int_0^\infty F_k(u, x)\theta(u)du = \frac{1}{2}[\phi(x+) + \phi(x-)],$$

and the proof of Theorem 9.3 is complete.

COROLLARY 9.3.1. *The hypothesis (9.1) may be replaced by the requirement that the integrals*

$$\int_{0+}^1 u^m \phi(u) du, \quad \int_1^{\infty} u^{-n} \phi(u) du$$

*exist, for some integers  $m, n \geq 0$ .*

We have only to verify (9.1). For  $u > 1$ , set

$$\psi(u) = \int_1^u t^{-n} \phi(t) dt.$$

Then

$$\begin{aligned} \gamma(u) &= \int_1^u \phi(t) dt = \int_1^u t^n d\psi(t) \\ &= \psi(u)u^n - n \int_1^u t^{n-1} \psi(t) dt = O(u^n), \quad u \rightarrow \infty. \end{aligned}$$

A similar argument applies when  $u \rightarrow 0$ .

10. **Inversion formulas.** We consider first the integral

$$(10.1) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{\phi(u) du}{t+u},$$

where  $\phi(u)$  is integrable on every  $(\epsilon, R)$ ,  $(0 < \epsilon < R < \infty)$ .

THEOREM 10.1. *If  $f(x)$  has the form (10.1), then*

$$(10.2) \quad \phi(x) = \lim_{k \rightarrow \infty} H_{k,x}[f(x)]$$

*for almost all  $x > 0$ .*

Because of Theorem 5.2, this inversion formula serves also for  $S_2$  transforms of the form

$$f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} \phi(t) dt.$$

By (4.3) the integral

$$\int_{0+}^u \phi(t) dt, \quad u > 0,$$

exists; by (4.4)

$$\int_{0+}^u \phi(t) dt = o(u), \quad u \rightarrow \infty.$$

Hence  $\phi(t)$  satisfies the hypotheses of Corollary 9.3.1, and

$$\lim_{k \rightarrow \infty} \int_{0+}^{\infty} F_k(u, x) \phi(u) du = \phi(x)$$

for almost all  $x$ . But by Theorem 6.1

$$H_{k,x}[f(x)] = \int_{0+}^{\infty} F_k(u, x) \phi(u) du, \quad k \geq 2.$$

The same reasoning leads to the following corollary:

**COROLLARY 10.1.1.** *If  $f(x)$  has the form (10.1), then (10.2) is true whenever  $\phi(x) = [\phi(x+) + \phi(x-)]/2$ .*

**THEOREM 10.2.** *If  $f(x)$  is an iterated Stieltjes transform of the form*

$$(10.3) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u},$$

with  $\alpha(t)$  a normalized function, of bounded variation on every  $(\epsilon, R)$ ,  $(0 < \epsilon < R < \infty)$ , then

$$(10.4) \quad \alpha(x) - \alpha(0+) = \lim_{k \rightarrow \infty} \int_{0+}^x H_{k,t}[f(x)] dt, \quad x > 0.$$

Because of Theorem 5.2, this inversion formula serves also for the  $S_2$  transform

$$f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t).$$

We begin by showing that the integral in (10.4) is defined. We write  $\beta(u) = \alpha(u) - \alpha(0+)$ , ( $u \geq 0$ ), and consider, for  $y > \epsilon > 0$  and  $k \geq 2$ , the integral

$$(10.5) \quad \begin{aligned} \int_{\epsilon}^y H_{k,x}[f(x)] dx &= \int_{\epsilon}^y dx \int_{0+}^{\infty} F_k(u, x) d\beta(u) \\ &= - \int_{\epsilon}^y dx \int_0^{\infty} \beta(u) F_k(u, x) du; \end{aligned}$$

the integrated terms in the integration by parts vanish because of (4.3), (4.4), and Lemma 7.3. We may change the order of integration in equation (10.5) and obtain

$$(10.6) \quad \int_{\epsilon}^y H_{k,x}[f(x)] dx = - \int_0^{\infty} \beta(u) du \int_{\epsilon}^y \frac{\partial}{\partial u} F_k(u, x) dx$$

if

$$(10.7) \quad \int_0^y dx \int_0^\infty \left| \beta(u) \frac{\partial}{\partial u} F_k(u, x) \right| du$$

is finite. By (4.3), (4.4), there is a constant  $A$  such that

$$|\beta(u)| \leq A(u+1), \quad 0 \leq u < \infty;$$

then, by use of (7.6) and Lemma 7.2,

$$\begin{aligned} \int_0^\infty \left| \beta(u) \frac{\partial}{\partial u} F_k(u, x) \right| du &\leq kA \int_0^\infty \frac{u+1}{u} F_k(u, x) du \\ &= kA \left\{ 1 + \frac{1}{x} \left( \frac{k-1}{k} \right)^2 \right\}. \end{aligned}$$

Hence the integral (10.7) is finite, and (10.6) is true. But  $xF_k(u, x)$  is homogeneous of order zero, so that by Euler's theorem

$$\frac{\partial}{\partial u} F_k(u, x) = -\frac{1}{u} \frac{\partial}{\partial x} [xF_k(u, x)],$$

and (10.6) becomes

$$(10.8) \quad \int_0^y H_{k,z}[f(x)]dx = \int_0^\infty u^{-1}\beta(u)[yF_k(u, y) - \epsilon F_k(u, \epsilon)]du.$$

Write

$$I(\epsilon) = \epsilon \int_0^\infty u^{-1}\beta(u)F_k(u, \epsilon)du.$$

For  $k \geq 3$ , at least,  $I(\epsilon)$  is defined; we shall show that  $\lim_{\epsilon \rightarrow 0} I(\epsilon) = 0$ . We can find, given  $\delta > 0$ , constants  $\eta$  and  $A > 0$  such that

$$\begin{aligned} |\beta(u)| &< \frac{1}{2}\delta, & 0 \leq u \leq \eta, \\ |\beta(u)| &< Au, & u \geq \eta. \end{aligned}$$

Then

$$I(\epsilon) \leq \frac{1}{2}\epsilon\delta \int_0^\infty u^{-1}F_k(u, \epsilon)du + A\epsilon \int_0^\infty F_k(u, \epsilon)du = \frac{\delta}{2} \left( \frac{k-1}{k} \right)^2 + A\epsilon.$$

For  $\epsilon < \delta/(2A)$ , we have  $|I(\epsilon)| < \delta$ ; and  $\delta$  was arbitrary. Hence we may let  $\epsilon \rightarrow 0$  in (10.8) and obtain

$$\int_{0+}^y H_{k,z}[f(x)]dx = y \int_0^\infty u^{-1}\beta(u)F_k(u, y)du, \quad y > 0, k \geq 3.$$

The function  $u^{-1}\beta(u)$  satisfies the hypotheses of Theorem 9.3, and satisfies (9.5) for every positive  $u$ . Therefore

$$\lim_{k \rightarrow \infty} \int_{0+}^y H_{k,z}[f(x)]dx = \beta(y) = \alpha(y) - \alpha(0+), \quad y > 0.$$

**11. The saltus operator.** We make the following definition:

**DEFINITION 11.1.** An operator  $h_{k,z}[f(x)]$  is defined by

$$h_{k,z}[f(x)] = x(8\pi/k)^{1/2}H_{k,z}[f(x)].$$

**THEOREM 11.1.** Under the hypotheses of Theorem 10.2,

$$(11.1) \quad \lim_{k \rightarrow \infty} h_{k,z}[f(x)] = \alpha(x+) - \alpha(x-), \quad x > 0.$$

We consider first the point  $x=1$ . Introduce the functions

$$\psi(u) = \begin{cases} \alpha(1-), & u < 1, \\ \alpha(1), & u = 1, \\ \alpha(1+), & u > 1, \end{cases}$$

$$\omega(u) = \alpha(u) - \psi(u);$$

$\omega(u)$  is continuous at  $u=1$ ,  $\omega(1)=0$ . Clearly

$$H_{k,1}[f(x)] = \int_{0+}^{\infty} F_k(u, 1)d\omega(u) + \int_0^{\infty} F_k(u, 1)d\psi(u).$$

Now

$$(11.2) \quad \int_0^{\infty} F_k(u, 1)d\psi(u) = [\alpha(1+) - \alpha(1-)]F_k(1, 1)$$

$$\sim [\alpha(1+) - \alpha(1-)]\left(\frac{k}{8\pi}\right)^{1/2}, \quad k \rightarrow \infty,$$

by Lemma 8.4. For  $k \geq 3$ ,

$$\int_{0+}^{\infty} F_k(u, 1)d\omega(u) = -\left(\int_0^{1-\eta} + \int_{1-\eta}^{1+\eta} + \int_{1+\eta}^{\infty}\right)\omega(u)dF_k(u, 1) = I_1 + I_2 + I_3,$$

where  $\eta$ , ( $0 < \eta < 1$ ), is chosen so that  $|\omega(u)| < \epsilon/2$  on  $(1-\eta, 1+\eta)$ ,  $\epsilon > 0$  being arbitrary. Now  $\omega(u)$  satisfies the hypotheses of Lemmas 9.1, 9.2, so that we obtain

$$\lim_{k \rightarrow \infty} (I_1 + I_3) = 0.$$

By Lemma 7.4,  $F_k(u, 1)$  increases on  $(1 - \eta, 1)$  and decreases on  $(1, 1 + \eta)$ , so that

$$\begin{aligned} |I_2| &\leq \int_{1-\eta}^1 |\omega(u)| dF_k(u, 1) - \int_1^{1+\eta} |\omega(u)| dF_k(u, 1) \\ &\leq \frac{\epsilon}{2} [2F_k(1, 1) - F_k(1 - \eta, 1) - F_k(1 + \eta, 1)] \sim \epsilon \left(\frac{k}{8\pi}\right)^{1/2}, \quad k \rightarrow \infty, \end{aligned}$$

by Lemmas 8.3 and 8.4. Thus

$$\limsup_{k \rightarrow \infty} \left(\frac{8\pi}{k}\right)^{1/2} \left| \int_{0+}^{\infty} F_k(u, 1) d\omega(u) \right| \leq \epsilon,$$

and  $\epsilon$  was arbitrary. Using also (11.2), we obtain (11.1) for  $x=1$ .

To establish (11.1) for  $x=x_0 \neq 1$ , we set  $x=x_0y$ ,  $t=x_0s$ ,  $u=x_0v$  in (10.3); a simple computation gives

$$g(y) = x_0 f(x_0 y) = \int_{0+}^{\infty} \frac{ds}{y+s} \int_{0+}^{\infty} \frac{d\beta(v)}{s+v}, \quad \beta(v) = \alpha(x_0 v).$$

By what has already been proved,

$$H_{k,1}[g(y)] \sim \left(\frac{k}{8\pi}\right)^{1/2} [\alpha(x_0+) - \alpha(x_0-)], \quad k \rightarrow \infty.$$

But

$$\begin{aligned} H_{k,1}[g(y)] &= \int_{0+}^{\infty} F_k(u, 1) d\alpha(x_0 u) = \int_{0+}^{\infty} F_k(x_0^{-1}u, 1) d\alpha(u) \\ &= x_0 \int_{0+}^{\infty} F_k(u, x_0) d\alpha(u) = x_0 H_{k,x_0}[f(x)]. \end{aligned}$$

Hence (11.1) is established in general.

### CHAPTER III. REPRESENTATION OF FUNCTIONS BY ITERATED STIELTJES TRANSFORMS

**12. Theorems on linear differential operators.** We consider operators of the form

$$(12.1) \quad L[f(x)] = \sum_{i=0}^n p_{n-i}(x) f^{(i)}(x),$$

where  $p_{n-i}(x) = B_{n-i}x^i$ ; the  $B_i$  are constants, and  $B_0 \neq 0$ . We shall call an operator of the form (12.1) an *Euler operator\** of order  $n$ . In this section we

\* Because  $L[f(x)] = g(x)$  is an "Euler differential equation." E. L. Ince, *Ordinary Differential Equations*, 1927, p. 141.

collect the properties of Euler operators which we shall need later.

**THEOREM 12.1.** *If  $L[f(x)]$  is an Euler operator of order  $n$ , there exists an operator  $\bar{L}[f(x)]$ , of order  $n$ , called the adjoint of  $L[f(x)]$ , such that for any functions  $f(x)$  and  $g(x)$  of class  $C^n$ ,*

$$(12.2) \quad g(x)L[f(x)] - f(x)\bar{L}[g(x)] = \frac{d}{dx}P[f(x), g(x)],$$

where

$$(12.3) \quad P[f(x), g(x)] = \sum_{i=1}^n \sum_{j=0}^{n-i} (-1)^i f^{(n-i-j)}(x) [p_{i-1}(x)g(x)]^{(j)}$$

$$(12.4) \quad = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} A_{ij} f^{(i)}(x) g^{(j)}(x) x^{i+j+1},$$

where the  $A_{ij}$  are constants. Moreover,

$$(12.5) \quad \bar{L}[f(x)] = \sum_{i=0}^n (-1)^i [p_{n-i}(x)f(x)]^{(i)}$$

$$(12.6) \quad = \sum_{i=0}^n A_i f^{(i)}(x) x^i, \quad A_n \neq 0,$$

where the  $A_i$  are constants; in particular,  $\bar{L}[f(x)]$  is an Euler operator.

The formulas (12.3) and (12.5) are the standard expressions.\* To reduce them to (12.4) and (12.6), respectively, one carries out the indicated differentiations and collects terms. The details are left to the reader.

**THEOREM 12.2.** *If  $L[f(x)]$  is an Euler operator of order  $n$ , if  $f(x)$  and  $g(x)$  are of class  $C^n$ , and if*

$$(12.7) \quad f^p(x)g^q(x)x^{p+q+1} \rightarrow 0, \quad 0 \leq p \leq n-1; 0 \leq q \leq n-1,$$

as  $x \rightarrow 0$  and as  $x \rightarrow \infty$ , then

$$\int_{0+}^{\infty} g(x)L[f(x)]dx = \int_{0+}^{\infty} f(x)\bar{L}[g(x)]dx$$

if either integral converges.

Because of (12.2),

$$\int_{\epsilon}^R g(x)L[f(x)]dx - \int_{\epsilon}^R f(x)\bar{L}[g(x)]dx = P[f(x), g(x)] \Big|_{\epsilon}^R, \quad 0 < \epsilon < R < \infty.$$

\* See, for example, E. L. Ince, op. cit., pp. 123-124.

By (12.4) and (12.7),

$$P[f(x), g(x)] = o(1), \quad x \rightarrow 0, x \rightarrow \infty.$$

**THEOREM 12.3.** *If  $L[f(x)]$  is an Euler operator of order  $n$ , a fundamental set of solutions of the differential equation  $L[f(x)] = 0$  is*

$$(12.8) \quad x^{a_i}, x^{a_i} \log x, \dots, x^{a_i} (\log x)^{b_i-1}, \quad i = 1, 2, \dots, j,$$

where the  $a_i$  are complex constants, the  $b_i$  are positive integers, and  $\sum_{i=1}^j b_i = n$ . Conversely, any set of functions of this form determines (except for a constant multiple) an Euler operator of order  $n$ , for which the functions form a fundamental set of solutions.

This is essentially a restatement of known results.\* We have

$$(12.9) \quad L[f(x)] = \sum_{i=0}^n B_{n-i} x^i f^{(i)}(x).$$

If we set  $x = e^z$ , it is easily verified that

$$x^i f^{(i)}(x) = \left[ \sum_{j=0}^{i-1} (D - j) \right] f(e^z), \dagger$$

where  $D$  denotes  $d/dz$ , so that

$$L[f(x)] = \sum_{i=0}^n B_{n-i} \left[ \sum_{j=0}^{i-1} (D - j) \right] f(e^z) = \sum_{i=0}^n B'_{n-i} D^i f(e^z) = M[f(e^z)].$$

The linear differential equation with constant coefficients,  $M[g(z)] = 0$ , has, as is well known, a fundamental set of solutions

$$e^{a_i z}, z e^{a_i z}, \dots, z^{b_i-1} e^{a_i z}, \quad i = 1, 2, \dots, j,$$

with

$$\sum_{i=1}^j b_i = n,$$

where the  $a_i$  are roots, of respective multiplicities  $b_i$ , of the algebraic equation  $\sum_{i=0}^n B'_{n-i} t^i = 0$ . Replacing  $z$  by  $\log x$ , we obtain the functions (12.8).

Conversely, let the functions (12.8) be given. There is a polynomial  $P(t) = \sum_{i=0}^n B'_{n-i} t^i$  having as roots the  $a_i$  with multiplicities  $b_i$ . We can write  $P(t)$  in the form

$$(12.10) \quad P(t) = \sum_{i=0}^n B_{n-i} \sum_{j=0}^{i-1} (t - j).$$

\* E. L. Ince, op. cit., pp. 141-142.

† An empty product denotes unity.

Then the operator  $L[f(x)]$ , defined by (12.9), with the constants  $B_i$  of (12.10), will have the functions (12.8) as a fundamental set of solutions, as the first part of the proof shows. Since  $P(t)$  is, except for a constant multiple, uniquely defined,  $L[f(x)]$  has the same property.

**THEOREM 12.4.** *If  $L[f(x)]$  is an Euler operator of order  $n$ , and  $f(x, y)$  is of class  $C^n$  and homogeneous of order  $-1$ , then  $L_x[f(x, y)] = \bar{L}_y[f(x, y)]$ .*

From the expressions (12.1) and (12.5) for  $L[f(x)]$  and  $\bar{L}[f(x)]$ , we see that it is sufficient to establish the theorem for the special operator  $L[f(x)] = x^k f^{(k)}(x)$ , ( $k=1, 2, \dots$ ), that is, to prove that

$$(12.11) \quad (-1)^k \frac{\partial^k}{\partial y^k} (y^k f(x, y)) = x^k \frac{\partial^k}{\partial x^k} f(x, y), \quad k = 1, 2, \dots, n.$$

For  $k=1$ , (12.11) is

$$-y \frac{\partial f}{\partial y} - f = x \frac{\partial f}{\partial x},$$

which is Euler's theorem for a homogeneous function of order  $-1$ .

We proceed by induction; assuming (12.11) for  $k-1$ , we establish it for  $k$ . The function  $x^{-k+1}y^kf$  is homogeneous of order zero; applying Euler's theorem, we have

$$\begin{aligned} \frac{\partial}{\partial y} \frac{y^k f}{x^{k-1}} &= -x \frac{\partial}{\partial x} \frac{y^{k-1} f}{x^{k-1}}, \\ \frac{\partial^k}{\partial y^k} \frac{y^k f}{x^{k-1}} &= -x \frac{\partial}{\partial x} \left[ \frac{1}{x^{k-1}} \frac{\partial^{k-1}}{\partial y^{k-1}} (y^{k-1} f) \right] \\ &= -x \frac{\partial}{\partial x} \left[ \frac{1}{x^{k-1}} (-1)^{k-1} x^{k-1} \frac{\partial^{k-1} f}{\partial x^{k-1}} \right] \\ &= (-1)^k x \frac{\partial^k f}{\partial x^k}; \end{aligned}$$

hence (12.11) is established for  $k$ .

**THEOREM 12.5.**  $H_{k,x}[f(x)]$  has as a set of fundamental solutions

$$(12.12) \quad x^n, x^n \log x, \quad n = -k, -k+1, \dots, -1, 0, 1, \dots, k-2.$$

There are  $4k-2$  of these functions; they are clearly linearly independent; it is easily verified that  $L_{k,x}[f(x)]$  annihilates  $x^n$ , and transforms  $x^n \log x$  into a constant multiple of  $x^n$ , ( $n = -k, -k+1, \dots, k-2$ ), so that  $H_{k,x}[f(x)]$  annihilates all the functions (12.12).

COROLLARY 12.5.1.  $H_{k,z}[f(x)]$  is an Euler operator.

13. An auxiliary kernel. We make the following definition:

DEFINITION 13.1. An operator  $Q[f(x)]$  is defined by

$$(13.1) \quad Q[f(x)] = x[x^2 f'(x)]'' = x^3 f'''(x) + 4x^2 f''(x) + 2xf'(x).$$

It is evident that  $Q[f(x)]$  is an Euler operator.

DEFINITION 13.2. A function  $E(x, t)$  is defined by

$$E(x, t) = Q_x \left[ \frac{\log(x/t)}{x-t} \right].$$

LEMMA 13.1. If  $x > 0, t > 0$ , then

$$(13.2) \quad E(x, t) = 2x \int_0^\infty \frac{u^2 du}{(x+u)^3(t+u)^2}.$$

We have

$$\begin{aligned} g(x) &\equiv \frac{\log(x/t)}{x-t} = \int_0^\infty \frac{du}{(x+u)(t+u)}, \\ g'(x) &= - \int_0^\infty \frac{du}{(x+u)^2(t+u)} = -\frac{1}{xt} + \int_0^\infty \frac{du}{(x+u)(t+u)^2}, \\ [x^2 g'(x)]'' &= 2 \int_0^\infty \frac{u^2 du}{(x+u)^3(t+u)^2}. \end{aligned}$$

LEMMA 13.2. If  $x > 0$  is fixed and  $n = 0, 1, 2, \dots$ , then

$$(13.3) \quad \frac{\partial^n}{\partial t^n} E(x, t) = O(t^{-n-2} \log t), \quad t \rightarrow \infty,$$

$$(13.4) \quad \frac{\partial^n}{\partial t^n} E(x, t) = O(t^{-n}), \quad t \rightarrow 0.$$

We have, from (13.2),

$$\begin{aligned} \frac{\partial^n}{\partial t^n} E(x, t) &= 2x(-1)^n(n+1)! \int_0^\infty \frac{u^2 du}{(x+u)^3(t+u)^{n+2}}, \\ (13.5) \quad t^n \left| \frac{\partial^n}{\partial t^n} E(x, t) \right| &\leq 2x(n+1)! \int_0^\infty \frac{u^2 du}{(x+u)^3(t+u)^2} \\ &= (n+1)! E(x, t). \end{aligned}$$

From (13.2), we see that (13.4) holds for  $n=0$ ; then (13.4) for  $n>0$  follows from (13.5). Also,

$$tE(x, t) \leq 2x \int_0^\infty \frac{du}{(x+u)(t+u)} = 2x \frac{\log(x/t)}{x-t},$$

and (13.3) for  $n=0$  follows; we obtain (13.3) for  $n>0$  by use of (13.5).

LEMMA 13.3. For  $k \geq 3$ ,

$$(13.6) \quad H_{k,z}[f(x)] = M_{k,z}\{Q[f(x)]\},$$

where  $M_{k,z}[f(x)]$  is an Euler operator of order  $4k-5$ .

Simple computation shows that, for any real  $n$ ,

$$Q[x^n] = n^2(n+1)x^n,$$

$$Q[x^n \log x] = n^2(n+1)x^n \log x + (3n^2 + 2n)x^n.$$

Hence  $Q[f(x)]$ , applied to the functions (12.12), gives either zero or a linear combination of the functions (12.12) other than 1,  $\log x$ , and  $x^{-1} \log x$ . By Theorem 12.3, there is an Euler operator  $M_{k,z}[f(x)]$ , of order  $4k-5$ , having the functions (12.12), other than 1,  $\log x$ , and  $x^{-1} \log x$ , as fundamental solutions. Then  $M_{k,z}\{Q[f(x)]\}$  annuls all the functions (12.12) and hence differs from  $H_{k,z}[f(x)]$  at most by a constant multiple. If this constant is suitably determined, (13.6) follows.

14. A general representation theorem. We prove first the following theorem:

THEOREM 14.1. If  $f(x)$  is of class  $C^\infty$ , if

$$(14.1) \quad f(x) = o(1), \quad x \rightarrow \infty,$$

$$(14.2) \quad f(x) = o(x^{-1}), \quad x \rightarrow 0,$$

and if

$$(14.3) \quad \int_1^x H_{k,t}[f(x)]dt < O(x), \quad x \rightarrow \infty,$$

$$(14.4) \quad \int_x^1 t^{-4k+2} H_{k,t}[f(x)]dt < O(x^{-4k+2}), \quad x \rightarrow 0,$$

each of (14.3) and (14.4) holding for an infinite sequence of positive integers  $k$ , then

$$(14.5) \quad Q[f(x)] = \lim_{k \rightarrow \infty} \int_{0+}^\infty H_{k,t}[f(x)]E(x, t)dt, \quad x > 0,$$

and, for  $n=0, 1, 2, \dots$ ,

$$(14.6) \quad f^{(n)}(x) = o(x^{-n}), \quad x \rightarrow \infty,$$

$$(14.7) \quad f^{(n)}(x) = o(x^{-n-1}), \quad x \rightarrow 0.$$

Let  $k$  be an integer greater than unity for which (14.3) holds. Since  $H_{k,z}[f(x)]$  is an Euler operator, we may apply a result of R. P. Boas,\* in virtue of which (14.1) and (14.3) imply (14.6) for  $n=1, 2, \dots, 4k-4$ . Since (14.3) holds for infinitely many  $k$ , (14.6) holds for all  $n$ . Similarly, (14.4) and (14.2) imply (14.7) for  $n=1, 2, \dots$ . Since  $f(x)$  satisfies (14.6), (14.7), the function  $Q[f(x)]$  also satisfies (14.6), (14.7) (see (13.1)). With Lemma 13.2, these relations imply

$$t^{p+q+1} \frac{d^p}{dt^p} Q[f(t)] \frac{\partial^q}{\partial t^q} E(x, t) = o(1)$$

for  $t \rightarrow 0, t \rightarrow \infty, p \geq 0, q \geq 0$ . Then by Theorem 12.2,

$$(14.8) \quad \int_{0+}^{\infty} E(x, t) M_{k,t} \{Q[f(t)]\} dt = \int_{0+}^{\infty} Q[f(t)] \bar{M}_{k,t} [E(x, t)] dt,$$

if either integral converges. But  $E(x, t)$  is homogeneous of order  $-1$ ; by Theorem 12.4, Definition 13.2, Lemma 13.3, and Corollary 6.1.1, we have

$$\begin{aligned} \bar{M}_{k,t} [E(x, t)] &= M_{k,z} [E(x, t)] \\ &= M_{k,z} \left\{ Q_z \left[ \frac{\log (x/t)}{x-t} \right] \right\} \\ &= H_{k,z} \left[ \frac{\log (x/t)}{x-t} \right] \\ &= F_k(t, x), \end{aligned} \quad k \geq 2.$$

Hence (14.8) becomes

$$(14.9) \quad \int_{0+}^{\infty} E(x, t) H_{k,t} [f(x)] dt = \int_0^{\infty} Q[f(t)] F_k(t, x) dt,$$

where the right-hand integral converges for  $k \geq 3$ , by Lemma 7.3 and the inequalities satisfied by  $Q[f(t)]$ . But  $Q[f(t)]$  satisfies the conditions of Theorem 9.3 and is continuous for  $t > 0$ ; hence

$$\lim_{k \rightarrow \infty} \int_0^{\infty} Q[f(t)] F_k(t, x) dt = Q[f(x)], \quad x > 0.$$

With (14.9) this yields (14.5).

\* R. P. Boas, *Asymptotic relations for derivatives*, Duke Mathematical Journal, vol. 3 (1937), pp. 637-646, Theorem 2, with  $\phi(x) = x\theta(x) = x$ , and Theorem 3, with  $\phi(x) = x^{-1}\theta(x) = x^{-1}$ .

**15. The iterated Stieltjes transform with non-decreasing determining function.** We make the following definition:

**DEFINITION 15.1.** A function  $f(x)$  will be said to satisfy Conditions A if and only if

- (i)  $f(x)$  is of class  $C^\infty$  on  $(0, \infty)$ ;
- (ii)  $f(x) = o(1)$ ,  $(x \rightarrow \infty)$ ;  $f(x) = o(x^{-1})$ ,  $(x \rightarrow 0)$ ;
- (iii) for an infinite sequence of positive integers  $k$ ,

$$H_{k,z}[f(x)] \geq 0, \quad 0 < x < \infty.$$

**THEOREM 15.1.** Conditions A are necessary and sufficient for  $f(x)$  to have the representation

$$(15.1) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

with  $\alpha(u)$  normalized and non-decreasing.

That Conditions A, (i) and A, (ii) are satisfied if  $f(x)$  has the form (15.1), we know by Theorem 4.1; Condition A, (iii) follows from

$$H_{k,z}[f(x)] = \int_{0+}^{\infty} F_k(u, x) d\alpha(u), \quad k \geq 2,$$

since  $F_k(u, x) \geq 0$ .

If  $f(x)$  satisfies Conditions A,  $-f(x)$  satisfies the hypotheses of Theorem 14.1, so that

$$(15.2) \quad Q[f(x)] = \lim_{k \rightarrow \infty} \int_{0+}^{\infty} H_{k,t}[f(x)] E(x, t) dt, \quad x > 0.$$

Formula (13.2) shows that  $E(x, t)$  is a positive decreasing function of  $t$ ,  $(0 < t < \infty)$ ; and  $E(x, x) = 1/(6x)$  (by Lemma 7.1). We then have, since  $H_{k,t}[f(x)] \geq 0$ ,

$$\begin{aligned} \int_{0+}^{\infty} H_{k,t}[f(x)] E(x, t) dt &\geq \int_{0+}^x H_{k,t}[f(x)] E(x, t) dt \geq E(x, x) \int_{0+}^x H_{k,t}[f(x)] dt, \\ \alpha_k(x) &= \int_{0+}^x H_{k,u}[f(x)] du \leq 6x \int_{0+}^{\infty} H_{k,t}[f(x)] E(x, t) dt. \end{aligned}$$

We define  $\alpha_k(0) = 0$ . If we refer to the proof of Theorem 14.1, we then have, by relation (14.9),

$$(15.3) \quad \alpha_k(x) \leq 6x \int_0^{\infty} Q[f(t)] F_k(t, x) dt, \quad x > 0.$$

But by (14.6), (14.7), there is a constant  $A$  such that

$$Q[f(t)] \leq A(1 + t^{-1}), \quad 0 < t < \infty;$$

hence, using (15.3) and Lemma 7.2, we have

$$\begin{aligned} \alpha_k(x) &\leq 6Ax \int_0^\infty (1 + t^{-1})F_k(t, x)dt \\ &= 6A \left\{ x + \left( \frac{k-1}{k} \right)^2 \right\}; \end{aligned}$$

and

$$(15.4) \quad \alpha_k(x) < 6A(x + 1), \quad 0 < x < \infty,$$

where  $A$  is independent of  $k$ .

The functions  $\alpha_k(x)$  (for  $k$  is the sequence of Conditions A) are non-decreasing and are bounded, uniformly with respect to  $k$ , in each interval  $(0, n)$ , ( $n = 1, 2, \dots$ ). By a theorem of E. Helly,\* we can select a subsequence converging in  $(0, 1)$ , a further subsequence converging in  $(0, 2)$ , and so on; by use of the diagonal process, we then obtain a subsequence  $\{\alpha_{k_t}(x)\}$ , converging in  $(0, \infty)$  to a non-decreasing function  $\alpha(x)$ . The relation in (15.2) states that

$$Q[f(x)] = \lim_{k \rightarrow \infty} \int_{0+}^\infty E(x, t) d\alpha_k(t) = \lim_{k \rightarrow \infty} \int_0^\infty E(x, t) d\alpha_k(t).$$

By use of (13.4), (15.4), and the Helly-Bray theorem,† it follows easily that

$$Q[f(x)] = \int_0^\infty E(x, t) d\alpha(t),$$

or that

$$\begin{aligned} [x^2 f'(x)]'' &= 2 \int_0^\infty d\alpha(t) \int_0^\infty \frac{u^2 du}{(x+u)^3(t+u)^2} \\ (15.5) \quad &= 2 \int_0^\infty \frac{t\psi(t)dt}{(x+t)^3}, \\ \psi(t) &= t \int_0^\infty \frac{d\alpha(u)}{(t+u)^2}. \end{aligned}$$

The changes of order of integration, here and for the remainder of the proof,

\* E. Helly, *Über lineare Funktionaloperationen*, Sitzungsberichte der Akademie der Wissenschaften, Vienna, vol. 121 (1921), p. 265.

† See, for example, G. C. Evans, *The Logarithmic Potential. Discontinuous Dirichlet and Neumann Problems*, American Mathematical Society Colloquium Publications, vol. 6, New York, 1927, p. 15.

are legitimate because the integrands are positive and  $\alpha(t)$  is non-decreasing.\*

Since  $[x^2 f'(x)]' = o(1)$ , ( $x \rightarrow \infty$ ), we may integrate (15.5) on  $(x, \infty)$ , obtaining

$$\begin{aligned} [x^2 f'(x)]' &= -2 \int_x^\infty dy \int_0^\infty \frac{t\psi(t)dt}{(y+t)^3} \\ &= - \int_0^\infty \frac{t\psi(t)dt}{(x+t)^2}. \end{aligned}$$

Since  $x^2 f'(x) = o(1)$ , ( $x \rightarrow 0$ ), we may now integrate on  $(0+, x)$  obtaining

$$\begin{aligned} x^2 f'(x) &= - \int_{0+}^x dy \int_0^\infty \frac{t\psi(t)dt}{(y+t)^2} \\ &= -x \int_0^\infty \frac{\psi(t)dt}{x+t}. \end{aligned}$$

The convergence of this integral implies, by use of Lemma 2.3, the convergence of

$$\begin{aligned} \int_t^\infty u^{-1}\psi(u)du &= \int_t^\infty du \int_0^\infty \frac{d\alpha(v)}{(u+v)^2} \\ &= \int_0^\infty \frac{d\alpha(u)}{t+u}. \end{aligned}$$

Then we have

$$\begin{aligned} xf'(x) &= \int_0^\infty \frac{t}{x+t} d_t \left( \int_0^\infty \frac{d\alpha(u)}{t+u} \right) \\ &= \frac{-\alpha(0+)}{x} - x \int_0^\infty \frac{dt}{(x+t)^2} \int_0^\infty \frac{d\alpha(u)}{t+u}, \end{aligned}$$

since

$$\phi(t) = \int_0^\infty \frac{d\alpha(u)}{t+u}$$

has the properties  $\phi(\infty) = 0$ , and  $\phi(t) \sim \alpha(0+)/t$ , ( $t \rightarrow 0$ ).† Thus

$$f'(x) = \frac{-\alpha(0+)}{x^2} - \int_0^\infty \frac{dt}{(x+t)^2} \int_0^\infty \frac{d\alpha(u)}{t+u};$$

since  $f(\infty) = 0$ , we may integrate on  $(x, \infty)$ , obtaining

\* The theorem which we use here is the analogue for Stieltjes integrals of the Fubini theorem for Lebesgue integrals; see S. Saks, *Theory of the Integral*, Monografie Matematyczne, vol. 7, Warsaw, 1937, p. 77.

† D. V. Widder, paper cited in §3, p. 10. By the way in which  $\alpha(t)$  was defined, we have  $\alpha(0) = 0$ .

$$f(x) = \frac{\alpha(0+)}{x} + \int_{0+}^{\infty} \frac{dt}{x+t} \int_0^{\infty} \frac{d\alpha(u)}{t+u};$$

since  $f(x) = o(x^{-1})$ ,  $(x \rightarrow 0)$ , and since the integral has the same property,  $\alpha(0+) = \alpha(0) = 0$ , and the proof is complete. If  $\alpha(t)$  were not normalized, normalization would not affect the representation (15.1); actually, because of Theorem 10.2, we see that our construction yields a normalized function  $\alpha(t)$ .

THEOREM 15.2. *Conditions A, and the additional condition*

$$(15.6) \quad f(x) = O(x^{-1} \log x), \quad x \rightarrow \infty,$$

*are necessary and sufficient for  $f(x)$  to have the representation*

$$(15.7) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

*with  $\alpha(t)$  normalized, non-decreasing, and bounded, on  $(0, \infty)$ .*

If  $f(x)$  has the representation in question, Conditions A are satisfied because of Theorem 15.1. To establish (15.6), we change the order of integration in (15.7)\* and write

$$\begin{aligned} f(x) &= \left( \int_{0+}^1 + \int_1^{\infty} \right) \frac{\log(x/t)}{x-t} d\alpha(t), & x > 1, \\ &= f_1(x) + f_2(x). \end{aligned}$$

Then

$$\begin{aligned} f_1(x) &\leq \frac{1}{x-1} \int_{0+}^1 \log(x/t) d\alpha(t) \\ &= \frac{\log x}{x-1} [\alpha(1) - \alpha(0+)] - \frac{1}{x-1} \int_{0+}^1 \log t d\alpha(t) \\ &= O(x^{-1} \log x), & x \rightarrow \infty, \\ f_2(x) &\leq \frac{\log x}{x-1} \int_1^{\infty} d\alpha(t) = O(x^{-1} \log x), & x \rightarrow \infty. \end{aligned}$$

Conversely, if  $f(x)$  satisfies Conditions A,  $f(x)$  has the representation (15.7), and it remains to show that (15.6) implies that  $\alpha(t)$  is bounded. Now (15.6) implies that for some constant  $M$ ,

$$(15.8) \quad \limsup_{x \rightarrow \infty} \frac{xf(x)}{\log x} \leq M.$$

\* See the last footnote but one.

We may change the order of integration in (15.7), obtaining

$$f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t);$$

since  $\alpha(t)$  is non-decreasing and  $(\log x - \log t)/(x-t)$  is a positive decreasing function of  $t$ , we have, for any  $R > 0$ ,

$$\begin{aligned} f(x) &\geq \int_{0+}^R \frac{\log(x/t)}{x-t} d\alpha(t) \\ &\geq \frac{\log(x/R)}{x-R} \int_{0+}^R d\alpha(t) = [\alpha(R) - \alpha(0+)] \frac{\log(x/R)}{x-R}; \end{aligned}$$

hence

$$(15.9) \quad \frac{xf(x)}{\log x} \geq [\alpha(R) - \alpha(0+)] \frac{\log(x/R)}{\log x} \frac{x}{x-R}.$$

If  $\alpha(t)$  were unbounded, we could choose  $R$  so large that  $\alpha(R) - \alpha(0+) > 2M$  and then obtain from (15.9)

$$\liminf_{x \rightarrow \infty} \frac{xf(x)}{\log x} \geq 2M,$$

which would contradict (15.8). Hence  $\alpha(t)$  is bounded.

**16. A Tauberian theorem.** Representation theorems for the iterated Stieltjes transform with determining function in a class other than that of non-decreasing functions are less easily established than the theorems of §15; there are no available theorems on change of order of integration to carry us, in general, from

$$Q[f(x)] = \int_0^{\infty} E(x, t) d\alpha(t)$$

to

$$f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}.$$

We shall use, instead, certain consequences of a Tauberian theorem of Hardy and Littlewood,\* which we quote as a lemma.

**LEMMA 16.1.** *If  $\phi(t)$  is integrable on every  $(\epsilon, R)$ ,  $(0 < \epsilon < R < \infty)$ , if*

$$(16.1) \quad h(x) = \int_{0+}^{\infty} \frac{\phi(t) dt}{(x+t)^p} \sim \frac{H}{x^p}, \quad x \rightarrow \infty,$$

\* G. H. Hardy and J. E. Littlewood, *Notes on the theory of series (XI): on Tauberian theorems*, Proceedings of the London Mathematical Society, (2), vol. 30 (1930), pp. 23-37; 33.

with  $0 < \sigma \leq \rho$ ,  $H \neq 0$ , and if, for almost all  $t$  not less than some  $t_0$ ,

$$(16.2) \quad \phi(t) \geq -Kt^{\rho-\sigma-1}, \quad K > 0;$$

then

$$(16.3) \quad \Phi(t) = \int_{0+}^t \phi(u) du \sim \frac{H\Gamma(\rho)}{\Gamma(\sigma)\Gamma(\rho-\sigma+1)} t^{\rho-\sigma}, \quad t \rightarrow \infty.$$

The theorem remains true when  $H=0$  if  $\rho > \sigma$ , and (16.1), (16.3) are interpreted as  $h(x) = o(x^{-\sigma})$ ,  $\Phi(t) = o(t^{\rho-\sigma})$ , respectively.

We have modified the original statement of the theorem somewhat, but the modifications are unimportant in their effect.

We shall need also the theorem resulting from one case ( $\rho = \sigma = 1$ ) of Lemma 16.1 by the substitutions  $t = u^{-1}$ ,  $y = x^{-1}$ . For convenience, we state it as the following lemma:

LEMMA 16.2. If the Stieltjes transform

$$g(y) = \int_{0+}^{\infty} \frac{\psi(u) du}{y + u}$$

converges, if  $\psi(u) > -M$ , ( $M > 0$ ), for almost all  $u$ , ( $0 < u < \infty$ ), and if

$$\lim_{y \rightarrow 0+} g(y) = G, \quad G \neq 0,$$

then

$$\int_{0+}^{\infty} u^{-1} \psi(u) du = G.$$

We use Lemmas 16.1, 16.2 to establish the following theorem:

THEOREM 16.3. Let  $\psi(t)$  be integrable on every  $(\epsilon, R)$ , ( $0 < \epsilon < R < \infty$ ), let

$$(16.4) \quad \psi(t) = O(1), \quad t \rightarrow \infty,$$

$$(16.5) \quad \psi(t) = O(t^{-1}), \quad t \rightarrow 0,$$

and let

$$(16.6) \quad \lim_{t \rightarrow 0+} t \int_t^1 u^{-1} \psi(u) du$$

exist. Assume that  $f(x)$ , of class  $C^\infty$ , satisfies

$$(16.7) \quad f^{(n)}(x) = \begin{cases} o(x^{-n}), & x \rightarrow \infty, \\ o(x^{-n-1}), & x \rightarrow 0, \end{cases}$$

for  $n = 0, 1, 2, 3$ , and let

$$(16.8) \quad [x^2 f'(x)]'' = 2 \int_0^\infty \frac{t\psi(t)dt}{(x+t)^3}, \quad 0 < x < \infty.$$

Then

$$\theta(t) = \int_t^\infty u^{-1}\psi(u)du, \quad t > 0,$$

is defined, and

$$f(x) = \int_{0+}^\infty \frac{\theta(t)}{x+t} dt, \quad 0 < x < \infty.$$

We integrate (16.8) on  $(x, y)$ ,  $(0 < x < y)$ ; application of Lemma 2.3 shows that the integral converges uniformly on  $(x, y)$ ; therefore

$$(16.9) \quad \begin{aligned} [y^2 f'(y)]' - [x^2 f'(x)]' &= \int_0^\infty t\psi(t) [(x+t)^{-2} - (y+t)^{-2}] dt \\ &= (y-x) \int_0^\infty \frac{t\psi(t)(x+y+2t)dt}{(x+t)^2(y+t)^2}. \end{aligned}$$

As  $y \rightarrow \infty$ ,  $[y^2 f'(y)]' \rightarrow 0$ , by (16.7). By use of (16.4) it is easily shown that we have

$$x(y-x) \int_0^\infty \frac{t\psi(t)dt}{(x+t)^2(y+t)^2} \rightarrow 0, \quad y \rightarrow \infty.$$

By (16.7) and (16.8)

$$\int_0^\infty \frac{t\psi(t)dt}{(x+t)^3} = o(x^{-1}), \quad x \rightarrow \infty;$$

by (16.4),  $|t\psi(t)| < Mt$ ,  $(t > 1)$ , for some constant  $M$ . By Lemma 16.1, with  $\rho=3$ ,  $\sigma=1$ ,  $H=0$ ,

$$\gamma(t) \equiv \int_0^t u\psi(u)du = o(t^2), \quad t \rightarrow \infty.$$

Then

$$\begin{aligned} \sigma(t) &\equiv \int_0^t \frac{u^2}{(x+u)^2} \psi(u)du = \int_0^t \frac{u}{(x+u)^2} d\gamma(u) \\ &= \frac{t\gamma(t)}{(x+t)^2} - \int_0^t \gamma(u) \frac{x-u}{(x+u)^3} du \\ &= o(t), \end{aligned} \quad t \rightarrow \infty.$$

Hence

$$y \int_0^\infty \frac{t^2 \psi(t) dt}{(x+t)^2(y+t)^2} = y \int_0^\infty \frac{d\sigma(t)}{(y+t)^2} = 2y \int_0^\infty \frac{\sigma(t) dt}{(y+t)^3} = o(1), \quad y \rightarrow \infty. *$$

Collecting results, we find from (16.9) that

$$\int_0^\infty \frac{t\psi(t) dt}{(x+t)^2(y+t)^2} \sim -y^{-2} [x^2 f'(x)]', \quad y \rightarrow \infty;$$

by (16.4), for fixed  $x > 0$  and some constant  $M$ ,

$$\left| \frac{t\psi(t)}{(x+t)^2} \right| < \frac{M}{t}, \quad t > 1.$$

It follows easily from (16.8) that  $f(x)$  is analytic. The relation  $[x^2 f'(x)]' \equiv 0$  is impossible (except in the trivial case  $f(x) \equiv 0$ , which we exclude from further consideration) because (16.7) excludes all linear combinations (with constant coefficients, not all zero) of the fundamental solutions 1 and  $x^{-1}$  of  $[x^2 f'(x)]' = 0$ ; hence  $[x^2 f'(x)]' = 0$  at most on a set  $S$  of isolated points. For  $x$  not in  $S$ , we can apply Lemma 16.1, with  $\rho = \sigma = 2$ , obtaining

$$\int_0^\infty \frac{t\psi(t) dt}{(x+t)^2} = -[x^2 f'(x)]'.$$

This holds for  $x$  in  $S$  as well, by continuity.

We integrate this relation on  $(y, x)$ ,  $(0 < y < x)$ , and obtain

$$x^2 f'(x) - y^2 f'(y) = (y - x) \int_0^\infty \frac{t\psi(t) dt}{(x+t)(y+t)}.$$

As  $y \rightarrow 0$  we have, by (16.7),  $y^2 f'(y) = o(1)$  and thus

$$\int_0^\infty \frac{t\psi(t) dt}{(x+t)(y+t)} \rightarrow -xf'(x), \quad y \rightarrow 0;$$

also, by (16.4) and (16.5), for fixed  $x > 0$  and some constant  $M$ ,

$$\left| \frac{t\psi(t)}{x+t} \right| < M, \quad t > 0.$$

By Lemma 16.2

$$-xf'(x) = \int_{0+}^\infty \frac{\psi(t) dt}{x+t},$$

for all  $x > 0$  for which  $f'(x) \neq 0$ ; since  $f'(x) = 0$  at most at a set of isolated points, this relation holds, by continuity, for all  $x > 0$ .

\* We leave to the reader the proof of the simple Abelian theorem used here.

A simple application of Lemma 2.3 shows that  $\theta(t)$  is defined; we then have

$$\begin{aligned} -xf'(x) &= -\int_{0+}^{\infty} \frac{t}{x+t} d\theta(t) \\ &= \frac{A}{x} + x \int_{0+}^{\infty} \frac{\theta(t)}{(x+t)^2} dt, \end{aligned}$$

where  $A = \lim_{t \rightarrow 0+} t\theta(t)$  is defined because (16.6) exists. We now have

$$f'(x) = -\frac{A}{x^2} - \int_{0+}^{\infty} \frac{\theta(t)}{(x+t)^2} dt.$$

Integrating on  $(x, y)$ ,  $(0 < x < y)$ , we have

$$\begin{aligned} f(x) - f(y) &= A \left( \frac{1}{x} - \frac{1}{y} \right) + (y-x) \int_{0+}^{\infty} \frac{\theta(t) dt}{(x+t)(y+t)}, \\ \int_{0+}^{\infty} \frac{\theta(t) dt}{(x+t)(y+t)} &\sim \frac{1}{y} \left[ f(x) - \frac{A}{x} \right], \quad y \rightarrow \infty, \end{aligned}$$

and

$$\left| \frac{\theta(t)}{x+t} \right| \leq \frac{M}{t}, \quad 0 < t < \infty,$$

for some constant  $M$ . By Lemma 16.1 (with  $\rho = \sigma = 1$ ),

$$f(x) - \frac{A}{x} = \int_{0+}^{\infty} \frac{\theta(t)}{x+t} dt$$

for all  $x > 0$  for which  $f(x) - Ax^{-1} \neq 0$ , and hence (by continuity) for all  $x > 0$ , since (16.7) excludes  $f(x) = Ax^{-1}$ . But  $f(x) = o(x^{-1})$ ,  $(x \rightarrow 0)$ ; and by (4.5) we obtain

$$\int_{0+}^{\infty} \frac{\theta(t)}{x+t} dt = o(x^{-1}), \quad x \rightarrow 0.$$

Hence  $A = 0$ , and the proof is complete.

**17. Determining function of bounded variation on  $(0, \infty)$ .** We introduce the following definition:

**DEFINITION 17.1.** A function  $f(x)$  is said to satisfy Conditions B if and only if

- (i)  $f(x)$  is of class  $C^\infty$  on  $(0, \infty)$ ;
- (ii)  $f(x) = o(x^{-1})$ ,  $(x \rightarrow 0)$ ;  $f(x) = o(1)$ ,  $(x \rightarrow \infty)$ ;
- (iii) for an infinite sequence of positive integers  $k$ ,

$$\int_0^\infty |H_{k,t}[f(x)]| dt < M,$$

where  $M$  is independent of  $k$ .

THEOREM 17.1. *Conditions B are necessary and sufficient for  $f(x)$  to have the representation*

$$(17.1) \quad f(x) = \int_{0+}^\infty \frac{dt}{x+t} \int_{0+}^\infty \frac{d\alpha(u)}{t+u},$$

where  $\alpha(u)$  is a normalized function of bounded variation on  $(0, \infty)$ .

If  $f(x)$  has the form (17.1), Conditions B, (i) and B, (ii) are satisfied, by Theorem 4.1. As for Condition B, (iii), by Theorem 6.1

$$\begin{aligned} |H_{k,t}[f(x)]| &= \left| \int_0^\infty F_k(u, t) d\alpha(u) \right| \\ &\leq \int_0^\infty F_k(u, t) |d\alpha(u)|, \quad k \geq 2; \end{aligned}$$

and

$$\begin{aligned} \int_0^\infty |H_{k,t}[f(x)]| dt &\leq \int_0^\infty dt \int_0^\infty F_k(u, t) |d\alpha(u)| \\ &= \int_0^\infty |d\alpha(u)| \int_0^\infty F_k(u, t) dt^* \\ &= \int_0^\infty |d\alpha(u)| = M. \end{aligned}$$

To establish the converse we apply Theorem 14.1. Conditions (14.1), (14.2), and (14.3) are evidently satisfied. To establish (14.4) we write

$$(17.2) \quad \alpha_k(t) = \int_0^t H_{k,u}[f(x)] du, \quad t \geq 0;$$

then

$$\begin{aligned} \int_x^1 t^{-4k+2} H_{k,t}[f(x)] dt &= \int_x^1 t^{-4k+2} d\alpha_k(t) \\ &= \alpha_k(1) - \alpha_k(x) x^{-4k+2} + (4k-2) \int_x^1 \alpha_k(t) t^{-4k+1} dt \\ &= o(x^{-4k+2}), \quad x \rightarrow 0, \end{aligned}$$

\* We have again used the Stieltjes analogue of the Fubini theorem. See the third footnote in §15.

since  $\alpha_k(t) = o(1)$ , ( $t \rightarrow 0$ ); this holds for the sequence of integers  $k$  of Condition B, (iii). Theorem 14.1 now gives

$$(17.3) \quad Q[f(x)] = \lim_{k \rightarrow \infty} \int_0^{\infty} E(x, t) d\alpha_k(t),$$

where  $\alpha_k(t)$  is defined by (17.2), and we think of  $k$  as restricted to the sequence of Condition B, (iii); Theorem 14.1 also shows that relations (14.6), (14.7) are satisfied.

Condition B, (iii) states that the functions  $\alpha_k(t)$  have uniformly bounded variation on  $(0, \infty)$ . By Helly's theorem\* we can pick a subsequence  $\{\alpha_{k_i}(t)\}$  converging to a function  $\alpha(t)$  of bounded variation on  $(0, \infty)$ . The function  $E(x, t)$  is continuous on  $(0, \infty)$  and approaches zero as  $t \rightarrow \infty$ ; it follows easily from the Helly-Bray theorem† that we may take the limit under the integral sign, over the sequence  $\{k_i\}$ , in (17.3). That is,

$$\begin{aligned} Q[f(x)] &= \int_0^{\infty} E(x, t) d\alpha(t) \\ &= \int_{0+}^{\infty} E(x, t) d\alpha(t), \end{aligned}$$

and

$$\int_0^{\infty} |d\alpha(t)| \leq M.$$

Using the expressions (13.1), (13.2) for  $Q[f(x)]$  and  $E(x, t)$ , we obtain

$$\begin{aligned} (17.4) \quad [x^2 f''(x)]'' &= 2 \int_{0+}^{\infty} d\alpha(t) \int_0^{\infty} \frac{u^2 du}{(x+u)^3 (t+u)^2} \\ &= 2 \int_0^{\infty} \frac{t\psi(t) dt}{(x+t)^3}, \end{aligned}$$

$$(17.5) \quad \psi(t) = t \int_{0+}^{\infty} \frac{d\alpha(u)}{(t+u)^2} \ddagger$$

We now apply Theorem 16.3. Clearly (16.4) and (16.5) are satisfied; also

$$g(t) = \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

exists because  $\alpha(t)$  has bounded variation on  $(0, \infty)$ ,  $g(t) = o(t^{-1})$ , ( $t \rightarrow 0$ ), (by Theorem 4.1), and

\* E. Helly, loc. cit.

† G. C. Evans, loc. cit.

‡ For the change of order of integration, see the third footnote in §15.

$$\psi(t) = -tg'(t),$$

so that the limit (16.6) exists. Conditions (16.7) are contained in (14.6), (14.7), which we saw above to be consequences of Conditions B, (ii) and B, (iii), through Theorem 14.1. Then by Theorem 16.3,

$$\begin{aligned} f(x) &= \int_{0+}^{\infty} \frac{dt}{x+t} \int_t^{\infty} \frac{\psi(u)}{u} du \\ &= - \int_{0+}^{\infty} \frac{dt}{x+t} \int_t^{\infty} g'(u) du \\ &= \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u} \end{aligned}$$

(where we have used the fact that  $g(\infty)=0$ ). By the definition of  $\alpha(t)$ , we obtain  $\alpha(0)=0$ ; that  $\alpha(t)$  is normalized follows from Theorem 10.2 and (17.2), because of the way in which  $\alpha(t)$  was defined.

18. **Determining function the integral of a function of  $L^p$ , ( $p>1$ ).** We make the following definition:

DEFINITION 18.1. A function  $f(x)$  satisfies Conditions C if and only if

- (i)  $f(x)$  is of class  $C^\infty$  on  $(0, \infty)$ ;
- (ii)  $f(x) = o(1)$ ,  $(x \rightarrow \infty)$ ;  $f(x) = o(x^{-1})$ ,  $(x \rightarrow 0)$ ;
- (iii) for an infinite sequence of positive integers  $k$ ,

$$\int_0^\infty |H_{k,t}[f(x)]|^p dt < M, \quad p > 1,$$

where  $M$  is independent of  $k$ .

THEOREM 18.1. Conditions C are necessary and sufficient for  $f(x)$  to have the representation

$$(18.1) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{\phi(u) du}{t+u},$$

with  $\phi(u)$  of class  $L^p$  on  $(0, \infty)$ , ( $p>1$ ).

If  $f(x)$  has the form (18.1), Conditions C, (i) and C, (ii) are satisfied, by Theorem 4.1. As for Condition C, (iii),

$$H_{k,z}[f(x)] = \int_0^\infty F_k(u, x) \phi(u) du;$$

then, by use of Hölder's inequality and the Fubini theorem,

$$\begin{aligned}
\left(\frac{k-1}{k}\right)^2 \int_0^\infty |\phi(u)|^p du &= \int_0^\infty |\phi(u)|^p du \int_0^\infty F_k(u, x) dx \\
&= \int_0^\infty dx \int_0^\infty F_k(u, x) |\phi(u)|^p du \\
&= \int_0^\infty \left\{ \left( \int_0^\infty F_k(u, x) du \right)^{p/q} \right. \\
&\quad \cdot \left. \int_0^\infty F_k(u, x) |\phi(u)|^p du \right\} dx, \quad 1/p + 1/q = 1, \\
&\geq \int_0^\infty \left( \int_0^\infty F_k(u, x)^{1/q} F_k(u, x)^{1/p} |\phi(u)|^p du \right) dx \\
&= \int_0^\infty \left( \int_0^\infty F_k(u, x) |\phi(u)|^p du \right) dx \\
&\geq \int_0^\infty \left| \int_0^\infty F_k(u, x) \phi(u) du \right|^p dx \\
&= \int_0^\infty |H_{k,x}[f(x)]|^p dx, \quad k = 2, 3, \dots
\end{aligned}$$

We may take  $M = \int_0^\infty |\phi(u)|^p du$ .

To establish the converse, we apply Theorem 14.1. We need only verify (14.3) and (14.4). To do this, we have, for the sequence of integers of Condition C, (iii) by Hölder's inequality,

$$\begin{aligned}
\left| \int_1^x H_{k,t}[f(x)] dt \right| &\leq x^{1/q} \left( \int_0^x |H_{k,t}[f(x)]|^p dt \right)^{1/p} = o(x), \quad x \rightarrow \infty, \\
\left| \int_x^1 t^{-4k+2} H_{k,t}[f(x)] dt \right| &\leq x^{-4k+2+1/q} \left( \int_x^\infty |H_{k,t}[f(x)]|^p dt \right)^{1/p} \\
&= o(x^{-4k+2}), \quad x \rightarrow 0.
\end{aligned}$$

Then by Theorem 14.1,

$$Q[f(x)] = \lim_{k \rightarrow \infty} \int_0^\infty H_{k,t}[f(x)] E(x, t) dt,$$

with  $k$  in the sequence in question; and (14.6), (14.7) are satisfied. By the weak compactness of the space  $L^p$ ,\* there is a function  $\phi(t)$  of  $L^p$ , such that for every function  $\omega(t)$  of  $L^q$ ,  $(1/p + 1/q = 1)$ ,

\* S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 130. Banach gives the theorem in question only for a finite interval, but it is equally valid for the infinite interval.

$$\lim_{k \rightarrow \infty} \int_0^{\infty} H_{k,t}[f(x)]\omega(t)dt = \int_0^{\infty} \phi(t)\omega(t)dt.$$

It follows from Lemma 13.2 that  $E(x, t)$  belongs to every  $L^q$ , ( $q > 1$ ); hence

$$Q[f(x)] = \int_0^{\infty} \phi(t)E(x, t)dt,$$

with  $\phi(t)$  belonging to  $L^p(0, \infty)$ . That is,

$$\begin{aligned} [x^2 f'(x)]'' &= 2 \int_0^{\infty} \phi(t)dt \int_0^{\infty} \frac{u^2 du}{(x+u)^3(t+u)^2} \\ &= 2 \int_0^{\infty} \frac{t\psi(t)dt}{(x+t)^3}, \\ \psi(t) &= t \int_0^{\infty} \frac{\phi(u)du}{(t+u)^2}; \end{aligned}$$

the change of the order of integration is justified by Fubini's theorem.

We now apply Theorem 16.3. By Hölder's inequality

$$\begin{aligned} |\psi(t)| &\leq t \left( \int_0^{\infty} |\phi(u)|^p du \right)^{1/p} t^{-2+1/q}, \quad 1/p + 1/q = 1, \\ &= O(t^{-1+1/q}), \quad t \rightarrow 0, t \rightarrow \infty; \end{aligned}$$

thus (16.4) and (16.5) are satisfied, and the limit (16.6) exists. Conditions (16.7) are included in (14.6) and (14.7), which we have already established. Moreover,

$$g(t) = \int_{0+}^{\infty} \frac{\phi(u)du}{t+u}, \quad t > 0,$$

is seen to exist, by another application of Hölder's inequality;  $\psi(t) = -tg'(t)$ ; and  $g(\infty) = 0$ . By Theorem 16.3

$$\begin{aligned} f(x) &= \int_{0+}^{\infty} \frac{dt}{x+t} \int_t^{\infty} \frac{\psi(u)}{u} du \\ &= \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{\phi(u)du}{t+u}, \end{aligned}$$

and the proof is complete.

**19. Determining function the integral of a function of class  $L$ .** Theorem 18.1 fails when  $p=1$ , since for  $p=1$ , Conditions C reduce to Conditions B. To treat the case  $p=1$ , we introduce the following definition:

DEFINITION 19.1. A function  $f(x)$  satisfies Conditions D if and only if

- (i)  $f(x)$  is of class  $C^\infty$  on  $(0, \infty)$ ;
- (ii)  $f(x) = o(1)$ ,  $(x \rightarrow \infty)$ ;  $f(x) = o(x^{-1})$ ,  $(x \rightarrow 0)$ ;
- (iii) for some infinite sequence of positive integers  $k$ ,  $H_{k,t}[f(x)]$  belongs to  $L(0, \infty)$ , and for  $m$  and  $n$  in the sequence,

$$\lim_{m,n \rightarrow \infty} \int_0^\infty |H_{m,t}[f(x)] - H_{n,t}[f(x)]| dt = 0.$$

THEOREM 19.1. Conditions D are necessary and sufficient for  $f(x)$  to have the representation

$$(19.1) \quad f(x) = \int_{0+}^\infty \frac{dt}{x+t} \int_{0+}^\infty \frac{\phi(u) du}{t+u},$$

with  $\phi(u)$  integrable on  $(0, \infty)$ .

If  $f(x)$  has the form (19.1), Conditions D, (i) and D, (ii) are certainly satisfied. To verify Condition D, (iii), we note that for  $k \geq 2$

$$\begin{aligned} \left(\frac{k-1}{k}\right)^2 \int_0^\infty |\phi(u)| du &= \int_0^\infty |\phi(u)| du \int_0^\infty F_k(u, t) dt \\ &= \int_0^\infty dt \int_0^\infty |\phi(u)| F_k(u, t) du \\ &\geq \int_0^\infty |H_{k,t}[f(x)]| dt, \end{aligned}$$

so that  $H_{k,t}[f(x)]$  belongs to  $L(0, \infty)$ , ( $k \geq 2$ ). In addition

$$\begin{aligned} |H_{k,t}[f(x)] - \phi(t)| &\leq \int_0^\infty F_k(u, t) |\phi(u) - \phi(t)| du \\ &= \int_0^\infty F_k(u, 1) |\phi(ut) - \phi(t)| du; \\ \int_0^\infty |H_{k,t}[f(x)] - \phi(t)| dt &\leq \int_0^\infty dt \int_0^\infty F_k(u, 1) |\phi(ut) - \phi(t)| du, \end{aligned}$$

if the iterated integral converges. It will converge if

$$(19.2) \quad \int_0^\infty F_k(u, 1) g(u) du$$

converges, where

$$g(u) = \int_0^\infty |\phi(ut) - \phi(t)| dt.$$

But for some constant  $A$

$$g(u) < A(1 + u^{-1});$$

hence (19.2) converges ( $k \geq 3$ ). Furthermore,  $g(u)$  is continuous at  $u=1$ , and  $g(1)=0$ .\* Corresponding to an arbitrary  $\epsilon > 0$ , we determine  $\delta$ , ( $0 < \delta < 1$ ), so that

$$g(u) < \epsilon, \quad |u - 1| < \delta.$$

Then

$$\begin{aligned} \int_0^\infty |H_{k,t}[f(x)] - \phi(t)| dt &\leq \left( \int_0^{1-\delta} + \int_{1-\delta}^{1+\delta} + \int_{1+\delta}^\infty \right) F_k(u, 1)g(u)du \\ &= I_1 + I_2 + I_3; \end{aligned}$$

$$I_2 \leq \epsilon \int_0^\infty F_k(u, 1)du = \epsilon,$$

$$I_1 \leq A \int_0^{1-\delta} (1 + u^{-1})F_k(u, 1)du \leq 2A \int_0^{1-\delta} u^{-1}F_k(u, 1)du$$

$$= 2A \int_0^{1-\delta} F_k(1, u)du$$

$$= 2AH_k(1 - \delta) = o(1), \dagger \quad k \rightarrow \infty,$$

$$I_3 \leq A \int_{1+\delta}^\infty (1 + u^{-1})F_k(u, 1)du \leq 2A \int_{1+\delta}^\infty F_k(u, 1)du$$

$$= 2Ad_k^2 \int_{1+\delta}^\infty u^k du \int_0^\infty \frac{t^{2k-1}dt}{(u+t)^{2k}(1+t)^{2k}}$$

$$\leq 2Ad_k^2 \int_{1+\delta}^\infty u^{k-2} du \int_0^\infty \frac{t^{2k-3}dt}{(u+t)^{2k-2}(1+t)^{2k-2}}$$

$$= \frac{2Ad_k^2}{d_{k-1}^2} \left\{ \left( \frac{k-2}{k-1} \right)^2 - H_{k-1}(1 + \delta) \right\}$$

$$= o(1), \quad k \rightarrow \infty.$$

It follows that

$$\lim_{k \rightarrow \infty} \int_0^\infty |H_{k,t}[f(x)] - \phi(t)| dt = 0,$$

which implies Condition D, (iii).

\* D. V. Widder, *A classification of generating functions*, these Transactions, vol. 39 (1936), p. 267.

†  $H_k(1 - \delta)$  is the function of Lemma 8.2.

We now establish the sufficiency of our conditions. Condition D, (iii) implies\* the existence of a function  $\phi(t)$ , integrable on  $(0, \infty)$ , such that

$$\lim_{k \rightarrow \infty} \int_0^\infty |H_{k,t}[f(x)]| dt = \int_0^\infty |\phi(t)| dt,$$

$$\lim_{k \rightarrow \infty} \int_0^t H_{k,u}[f(x)] du = \int_0^t \phi(u) du, \quad t > 0,$$

where  $k$  runs through the sequence of Condition D, (iii). Consequently,

$$\int_0^\infty |H_{k,t}[f(x)]| dt \leq \int_0^\infty |\phi(t)| dt + 1$$

for  $k$  greater than some  $k_0$ , and  $k$  in the sequence. Thus  $f(x)$  satisfies Conditions B, and by Theorem 17.1

$$f(x) = \int_{0+}^\infty \frac{dt}{x+t} \int_{0+}^\infty \frac{d\alpha(u)}{t+u},$$

with  $\alpha(t)$  a normalized function, of bounded variation on  $(0, \infty)$ . By Theorem 10.2,

$$\alpha(t) - \alpha(0+) = \lim_{k \rightarrow \infty} \int_0^t H_{k,u}[f(x)] du, \quad t > 0.$$

But

$$\lim_{k \rightarrow \infty} \int_0^t H_{k,u}[f(x)] du = \int_0^t \phi(u) du,$$

where  $\{k_i\}$  is a certain subsequence of the integers, and consequently

$$\alpha(t) - \alpha(0+) = \int_0^t \phi(u) du.$$

Hence  $f(x)$  has the form (19.1).

**20. Determining function the integral of a bounded function.** We introduce the following definition:

**DEFINITION 20.1.** A function  $f(x)$  satisfies Conditions E if and only if

- (i)  $f(x)$  is of class  $C^\infty$  on  $(0, \infty)$ ;
- (ii)  $f(x) = o(1)$ ,  $(x \rightarrow \infty)$ ;  $f(x) = o(x^{-1})$ ,  $(x \rightarrow 0)$ ;
- (iii) for an infinite sequence of positive integers  $k$ ,

$$|H_{k,x}[f(x)]| \leq M, \quad 0 < x < \infty,$$

where  $M$  is independent of  $k$ .

\* See, for example, E. C. Titchmarsh, *The Theory of Functions*, 1932, pp. 387 ff.

THEOREM 20.1. *Conditions E are necessary and sufficient for  $f(x)$  to have the representation*

$$(20.1) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{\phi(u)du}{t+u},$$

with  $\phi(u)$  bounded almost everywhere.

If  $f(x)$  has the form (20.1), we have only to verify Condition E, (iii). For  $k \geq 2$  we have

$$|H_{k,z}[f(x)]| = \left| \int_0^{\infty} F_k(u, x) \phi(u) du \right| \leq M \int_0^{\infty} F_k(u, x) du = M,$$

where  $|\phi(u)| \leq M$  almost everywhere.

To show that the conditions are sufficient, we use Theorem 14.1, whose hypotheses are evidently fulfilled. It follows that

$$Q[f(x)] = \lim_{k \rightarrow \infty} \int_0^{\infty} H_{k,t}[f(x)] E(x, t) dt.$$

By the weak compactness of the space of functions bounded almost everywhere,\* there exists a function  $\phi(t)$ , bounded almost everywhere, such that for every function  $\omega(t)$  of  $L(0, \infty)$

$$\lim_{k \rightarrow \infty} \int_0^{\infty} H_{k,t}[f(x)] \omega(t) dt = \int_0^{\infty} \phi(t) \omega(t) dt.$$

Since  $E(x, t)$  belongs to  $L(0, \infty)$ , we obtain

$$\begin{aligned} [x^2 f'(x)]'' &= 2 \int_0^{\infty} \phi(t) dt \int_0^{\infty} \frac{u^2 du}{(x+u)^3 (t+u)^2} = 2 \int_0^{\infty} \frac{\psi(t) dt}{(x+t)^3}, \\ \psi(t) &= t \int_0^{\infty} \frac{\phi(u) du}{(t+u)^2}. \end{aligned}$$

We now apply Theorem 16.3. Since

$$|\psi(t)| \leq M,$$

the conditions of that theorem are satisfied, and

$$(20.2) \quad \begin{aligned} f(x) &= \int_{0+}^{\infty} \frac{\theta(t)}{x+t} dt, \\ \theta(x) &= \int_{\infty}^{\infty} dt \int_0^{\infty} \frac{\phi(u) du}{(t+u)^2}. \end{aligned}$$

\* S. Banach, loc. cit. Banach gives the theorem only for a finite interval.

For  $y > x$  we have

$$\theta(y) - \theta(x) = (x - y) \int_0^\infty \frac{\phi(u) du}{(x+u)(y+u)},$$

since the integral defining  $\psi(t)$  is evidently uniformly convergent ( $x \leq t \leq y$ ). Hence

$$\int_0^\infty \frac{\phi(u) du}{(x+u)(y+u)} \sim \frac{\theta(x)}{y} \quad y \rightarrow \infty;$$

and

$$\left| \frac{\phi(u)}{x+u} \right| \leq \frac{M}{u}$$

for almost all  $u$ , and for fixed  $x$ . By Lemma 16.1 (with  $\rho = \sigma = 1$ ),

$$(20.3) \quad \theta(t) = \int_0^\infty \frac{\phi(u) du}{t+u},$$

for all  $t > 0$  for which  $\theta(t) \neq 0$ . Since  $\theta(t) = 0$  at most on a set of isolated points (except in the trivial case  $f(x) \equiv 0$ ), (20.3) holds by continuity for all  $t > 0$ . Substitution of this formula for  $\theta(t)$  into (20.2) completes the proof.

#### CHAPTER IV. THE REPRESENTATION OF FUNCTIONS BY $S_2$ TRANSFORMS

**21. Determining function non-decreasing; determining function the integral of a function of class  $L^p$ , ( $p > 1$ ).** In these two cases the  $S_2$  transform and the iterated Stieltjes transform are equivalent. The  $S_2$  transform is obtained from the iterated Stieltjes transform by a formal change of the order of integration. If the determining function is non-decreasing this formal process is legitimate; it is also legitimate if the determining function is the integral of a function of class  $L^p$ , ( $p > 1$ ). In fact, if  $\phi(t)$  belongs to  $L^p$ , ( $p > 1$ ), then

$$\begin{aligned} \left| \int_0^\infty \frac{\phi(u) du}{t+u} \right| &\leq \left( \int_0^\infty |\phi(u)|^p du \right)^{1/p} \left( \int_0^\infty \frac{du}{(t+u)^q} \right)^{1/q}, \quad 1/p + 1/q = 1, \\ &= (q-1)^{-1/q} t^{-1/p} \left( \int_0^\infty |\phi(u)|^p du \right)^{1/p}, \end{aligned}$$

and since  $t^{-1/p}/(x+t)$  is integrable on  $(0, \infty)$ , ( $x > 0$ ),

$$\int_0^\infty \frac{dt}{x+t} \int_0^\infty \frac{|\phi(u)| du}{t+u}$$

exists and dominates

$$\int_0^\infty \frac{dt}{x+t} \int_0^\infty \frac{\phi(u)du}{t+u}.$$

We may therefore state the following theorems:

**THEOREM 21.1.** *A necessary and sufficient condition that*

$$f(x) = \int_{0+}^\infty \frac{\log(x/t)}{x-t} d\alpha(t),$$

*with  $\alpha(t)$  normalized and non-decreasing, is that  $f(x)$  satisfy Conditions A.*

**THEOREM 21.2.** *A necessary and sufficient condition that*

$$f(x) = \int_{0+}^\infty \frac{\log(x/t)}{x-t} \phi(t)dt,$$

*with  $\phi(t)$  belonging to  $L^p(0, \infty)$ , ( $p > 1$ ), is that  $f(x)$  satisfy Conditions C.*

Corresponding to the representation theorems for the iterated Stieltjes transform in the other cases, there are representation theorems for the  $S_2$  transform; in each case an auxiliary condition is imposed to make application of Theorem 5.3 possible.

**22. A lemma.** We can make the following statement:

**LEMMA 22.1.** *Let*

$$H_k(y) = \int_0^y F_k(1, x)dx.*$$

*Then there is a constant  $A$  such that*

$$(22.1) \quad y^{-1/2}H_k(y) \leq A, \quad 0 < y \leq 1/2,$$

*uniformly with respect to  $k$ , ( $k \geq 2$ ).*

We refer to the proof of Lemma 8.2 (page 23), where we find the relation

$$H_k(y) \leq 2d_k \int_0^{y^{1/2}} \frac{t^{k-1}dt}{(t+1)^{2k}}.$$

Since  $t(t+1)^{-2}$  increases on  $(0, y^{1/2})$ ; we have

$$H_k(y) \leq 2d_k \left( \frac{y^{1/2}}{(y^{1/2}+1)^2} \right)^{k-1} \int_0^{y^{1/2}} \frac{dt}{(1+t)^2} = 2d_k \left( \frac{y^{1/2}}{(y^{1/2}+1)^2} \right)^{k-1} \frac{y^{1/2}}{1+y^{1/2}}.$$

But

$$d_k = \frac{(2k-1)!}{k!(k-2)!}, \quad k \geq 2.$$

\*  $H_k(y)$  is the function of Lemma 8.2.

By use of Stirling's formula, we see that there is a positive constant  $B$  such that

$$d_k \leq B k^{1/2} 2^{2k}, \quad k \geq 2.$$

Since  $0 < y \leq 1/2$ , there is a constant  $\lambda < 1$  such that

$$4y^{1/2}(y^{1/2} + 1)^{-2} \leq \lambda.$$

Then

$$H_k(y) \leq 8Bk^{1/2}\lambda^{k-1}y^{1/2}(1 + y^{1/2})^{-1} \leq Ay^{1/2}, \quad k \geq 2,$$

for a suitably chosen constant  $A$ .

**23. Determining function of bounded variation on  $(0, \infty)$ .** We prove the following theorem:

**THEOREM 23.1.** *A necessary and sufficient condition that  $f(x)$  have the representation*

$$(23.1) \quad f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t),$$

with  $\alpha(t)$  a normalized function, of bounded variation on  $(0, \infty)$ , is that  $f(x)$  should satisfy Conditions B, and that for an infinite sequence of positive integers  $k$ ,

$$(23.2) \quad \log \frac{1}{t} \left| \int_{0+}^t H_{k,u}[f(x)] du \right| < \epsilon(t), \quad 0 < t \leq 1/4,$$

where  $\lim_{t \rightarrow 0} \epsilon(t) = 0$ , and  $\epsilon(t)$  is independent of  $k$ .

We show first that if  $f(x)$  has the form (23.1), then (23.2) is satisfied. By Theorem 5.2,  $f(x)$  is an iterated Stieltjes transform; hence

$$(23.3) \quad \begin{aligned} H_{k,t}[f(x)] &= \int_0^{\infty} F_k(u, t) d\alpha(u), \quad k \geq 2, \\ \int_{0+}^x H_{k,t}[f(x)] dt &= \int_{0+}^x dt \int_0^{\infty} F_k(u, t) d\alpha(u) = \int_0^{\infty} d\alpha(u) \int_0^x F_k(u, t) dt; \end{aligned}$$

the change of the order of integration is legitimate because under our hypotheses, the last integral is absolutely convergent.\*

The function

$$\int_0^x F_k(u, t) dt$$

\* See the third footnote in §15.

is a decreasing function of  $u$ . For,  $tF_k(u, t)$  is homogeneous of order zero,

$$\frac{\partial}{\partial u} F_k(u, t) = -\frac{1}{u} \frac{\partial}{\partial t} (tF_k(u, t)),$$

and

$$\begin{aligned} \frac{\partial}{\partial u} \int_0^x F_k(u, t) dt &= \int_0^x \frac{\partial}{\partial u} F_k(u, t) dt \\ &= -xu^{-1}F_k(u, x) < 0, \quad u > 0, x > 0. \end{aligned}$$

We now have, from (23.3),

$$\int_{0+}^x H_{k,t}[f(x)] dt = \left( \int_{0+}^{x^{1/2}} + \int_{x^{1/2}}^{\infty} \right) d\alpha(u) \int_0^x F_k(u, t) dt = I_1 + I_2.$$

Using Lemma 2.2, which applies because  $\int_0^x F_k(u, t) dt$  is a positive decreasing function of  $u$ , we obtain

$$|I_1| \leq \left( \lim_{u \rightarrow 0+} \int_0^x F_k(u, t) dt \right) \text{u.b.}_{0 \leq v \leq x^{1/2}} |\alpha(y) - \alpha(0+)|.$$

But

$$\int_0^x F_k(u, t) dt \leq \int_0^{\infty} F_k(u, t) dt = \left( \frac{k-1}{k} \right)^2 < 1,$$

and

$$(23.4) \quad |I_1| \leq \text{u.b.}_{0 \leq v \leq x^{1/2}} |\alpha(y) - \alpha(0+)|.$$

Let

$$\int_0^{\infty} |d\alpha(u)| = M.$$

Since  $\int_0^x F_k(u, t) dt$  is a decreasing function of  $u$ ,

$$\begin{aligned} |I_2| &\leq M \int_0^x F_k(x^{1/2}, t) dt = Mx^{1/2} \int_0^{x^{1/2}} F_k(x^{1/2}, x^{1/2}u) du, \quad t = ux^{1/2}, \\ &= M \int_0^{x^{1/2}} F_k(1, u) du = MH_k(x^{1/2}). \end{aligned}$$

According to Lemma 22.1, then, there is a constant  $A$  such that

$$|I_2| \leq AMx^{1/4}, \quad k \geq 2, 0 < x \leq 1/4.$$

Combining this with (23.4), we have

$$|I_1 + I_2| \leq \epsilon(x), \quad 0 < x \leq 1/4,$$

where

$$\epsilon(x) = \text{u. b.}_{0 \leq y \leq x^{1/2}} |\alpha(y) - \alpha(0+)| + AMx^{1/4}.$$

The function  $\epsilon(x)$  is independent of  $k$ , ( $k \geq 2$ ); and  $\epsilon(x) = o(-1/\log x)$ , ( $x \rightarrow 0$ ), since by Theorem 3.1,

$$\text{u. b.}_{0 \leq y \leq x^{1/2}} |\alpha(y) - \alpha(0+)| = o(-1/\log x^{1/2}) = o(-1/\log x).$$

Conversely let us suppose that  $f(x)$  satisfies the conditions of the theorem. Conditions B imply that

$$(23.5) \quad f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u},$$

with  $\alpha(u)$  a normalized function of bounded variation on  $(0, \infty)$ . By Theorem 10.2,

$$\alpha(u) - \alpha(0+) = \lim_{k \rightarrow \infty} \int_{0+}^u H_{k,t}[f(x)]dt, \quad u > 0.$$

Therefore

$$|\alpha(u) - \alpha(0+)| \log(1/u) = \lim_{k \rightarrow \infty} \log(1/u) \left| \int_{0+}^u H_{k,t}[f(x)]dt \right|.$$

But by (23.2),

$$\log(1/u) \left| \int_0^u H_{k,t}[f(x)]dt \right| \leq \epsilon(u), \quad 0 < u \leq 1/4,$$

for some sequence of integers  $k$ . Hence

$$(23.6) \quad |\alpha(u) - \alpha(0+)| \log(1/u) \leq \epsilon(u), \quad 0 < u \leq 1/4.$$

Furthermore,

$$\int_0^{\infty} |d\alpha(u)| = M,$$

and for  $t > 0$ ,

$$(23.7) \quad \left| \int_t^{\infty} \frac{d\alpha(u)}{u} \right| \leq \frac{M}{t} = o(1/\log t), \quad t \rightarrow \infty.$$

Conditions (23.6) and (23.7) are the conditions of Theorem 5.2; since they are satisfied, we may change the order of integration in (23.5) to obtain the representation (23.1) for  $f(x)$ .

The condition (23.2) appears highly artificial; one might hope to replace it by a weaker condition which, together with Conditions A would still be sufficient for  $f(x)$  to have the representation (23.1). This, however, does not appear to be possible. Inequality (23.2) states that

$$\int_{0+}^x H_{k,t}[f(x)]dt = o(-1/\log x), \quad x \rightarrow 0,$$

where the function  $o(-1/\log x)$  is independent of  $k$ ; if the uniform  $o(-1/\log x)$  is replaced by a uniform  $O(-1/\log x)$  and a non-uniform  $o(-1/\log x)$ , then Theorem 23.1 ceases to be true. This is verified by the following theorem:

**THEOREM 23.2.** *There exists a normalized function  $\alpha(u)$ , of bounded variation on  $(0, \infty)$ , such that*

$$f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u}$$

*converges,*

$$\int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t)$$

*diverges,*

$$(23.8) \quad \log(1/x) \left| \int_{0+}^x H_{k,t}[f(x)]dt \right| \leq M, \quad k \geq 2; 0 < x \leq 1,$$

*with  $M$  independent of  $k$ , and*

$$(23.9) \quad \log(1/x) \int_{0+}^x H_{k,t}[f(x)]dt = o(1), \quad x \rightarrow 0,$$

*for each  $k \geq 2$ .*

The function  $\alpha(u)$  is defined by means of the sequences  $\{u_n\}$ ,  $\{u'_n\}$ , as in Theorem 5.4; the sequences are now further restricted by the conditions that

$$u'_{n+1} < u_n^2, \quad u'_n < 2u_n, \quad n = 1, 2, \dots,$$

and that

$$\sum_{n=1}^{\infty} (u'_n - u_n)u_n^{-2}$$

converges. For example, we might have

$$u_n = 2^{-2^{n^2}}, \quad u'_n = (1 + n^{-2}u_n)u_n.$$

The first two statements of the theorem were established in Theorem 5.4.

The proof of the necessity of the conditions of Theorem 23.1 shows that (23.8) is satisfied if

$$\alpha(u) = O(-1/\log u), \quad u \rightarrow 0,$$

(which is readily verified in this case), and that

$$\begin{aligned} \int_{0+}^x H_{k,t}[f(x)]dt &= \left( \int_{0+}^{x^{1/2}} + \int_{x^{1/2}}^{\infty} \right) d\alpha(u) \int_0^x F_k(u, t)dt \\ &= I_1 + I_2, \end{aligned}$$

with

$$I_2 = O(x^{1/4}) = o(-1/\log x), \quad x \rightarrow 0.$$

It remains only to show that for  $k \geq 2$ ,

$$(23.10) \quad I_1 = o(-1/\log x), \quad x \rightarrow 0.$$

Take  $x < u_1^2$ . We have  $I_1 = Q + R$ , where

$$Q = \sum_{n=n_0+1}^{\infty} \frac{-1}{\log u_n} \int_0^x [F_k(u_n, t) - F_k(u'_n, t)]dt,$$

and  $n_0$  is determined by the condition  $u_{n_0} \leq x^{1/2} < u_{n_0-1}$ ;

$$R = \begin{cases} - (2 \log u_{n_0})^{-1} \int_0^x F_k(u_{n_0}, t)dt, & x^{1/2} = u_{n_0}, \\ - (\log u_{n_0})^{-1} \int_0^x F_k(u_{n_0}, t)dt, & u_{n_0} < x^{1/2} < u'_{n_0}, \\ - (\log u_{n_0})^{-1} \int_0^x [F_k(u_{n_0}, t) - F_k(u'_{n_0}, t)]dt, & u'_{n_0} < x^{1/2} < u_{n_0} - 1, \\ - (2 \log u_{n_0})^{-1} \left\{ \int_0^x F_k(u_{n_0}, t)dt + \int_0^x [F_k(u_{n_0}, t) - F_k(u'_{n_0}, t)]dt \right\}, & x^{1/2} = u'_{n_0}. \end{cases}$$

Now

$$\frac{\partial}{\partial u} \int_0^x F_k(u, t)dt = - \frac{x}{u} F_k(u, x);^*$$

hence

$$\int_0^x [F_k(u_n, t) - F_k(u'_n, t)]dt = x(u'_n - u_n)(u''_n)^{-1} F_k(u''_n, x), \quad u'_n < u''_n < u_n,$$

\* See p. 32.

so that

$$(23.11) \quad Q = x \sum_{n=n_0+1}^{\infty} \frac{u_n - u'_n}{u_n'' \log u_n} F_k(u_n'', x).$$

For  $n \geq n_0 + 1$ , we have  $u'_n < u_{n_0} \leq x^{1/2}$ ; since  $u'_{n_0+1} < u_{n_0}^2$ , we have  $u'_{n_0+1} < x$ ; and (Lemma 7.4)  $F_k(u, x)$  is an increasing function of  $u$  for  $u \leq u'_{n_0+1}$ , ( $k \geq 2$ ). Thus, for  $n \geq n_0 + 1$ ,

$$F_k(u_n'', x) \leq F_k(u'_n, x), \quad k \geq 2.$$

But

$$\begin{aligned} F_k(u'_n, x) &= d_k^2 u_n'^k x^{k-1} \int_0^{\infty} \frac{s^{2k-1} ds}{(s + u_n')^{2k} (s + x)^{2k}} \\ &\leq d_k^2 u_n'^k \int_0^{\infty} \frac{s^{k-2} ds}{(s + u_n')^{2k}} \\ &= \frac{A_k}{u_n'} \leq \frac{A_k}{u_n}, \end{aligned} \quad n \geq n_0 + 1,$$

where  $A_k$  depends only on  $k$ . The function  $-1/\log u$  increases for  $0 < u < 1$ ; hence for  $n \geq n_0 + 1$ , one has

$$\frac{-1}{\log u_n} < \frac{-1}{\log u_{n_0}} \leq \frac{-1}{\log x^{1/2}}.$$

Relation (23.11) now gives

$$\begin{aligned} 0 < Q &\leq \frac{2A_k x}{\log x} \sum_{n=n_0+1}^{\infty} \frac{u_n - u'_n}{u_n^2} \leq \frac{2A_k x}{\log x} \sum_{n=1}^{\infty} \frac{u_n - u'_n}{u_n^2} \\ &= O(-x/\log x) = o(-1/\log x), \end{aligned} \quad x \rightarrow 0.$$

If  $u_{n_0} \leq x^{1/2} \leq u'_{n_0}$ , then

$$\begin{aligned} 0 &< \frac{-1}{\log u_{n_0}} \int_0^x F_k(u_{n_0}, t) dt \\ &= \frac{-2d_k^2 u_{n_0}^k}{\log x} \int_0^x t^{k-1} dt \int_0^{\infty} \frac{s^{2k-1} ds}{(s + t)^{2k} (s + u_{n_0})^{2k}} \\ &\leq \frac{-2d_k^2 u_{n_0}^k}{\log x} \int_0^x dt \int_0^{\infty} \frac{s^{k-2} ds}{(s + u_{n_0})^{2k}} \\ &= \frac{-A'_k x}{u_{n_0} \log x} \leq \frac{-2A'_k x}{u_{n_0}' \log x} \leq \frac{-2A'_k x^{1/2}}{\log x}, \end{aligned}$$

where  $A'_k$  depends only on  $k$ . Therefore,

$$0 < R \leq \frac{-2A'_k x^{1/2}}{\log x}, \quad u_{n_0} \leq x^{1/2} < u'_{n_0}.$$

If, on the other hand,  $u'_{n_0} \leq x^{1/2} < u_{n_0-1}$ , then

$$\begin{aligned} 0 &< \frac{-1}{\log u_{n_0}} \int_0^x [F_k(u_{n_0}, t) - F_k(u'_{n_0}, t)] dt \\ &= \frac{x(u_{n_0} - u'_{n_0})}{u'_{n_0} \log u_{n_0}} F_k(u''_{n_0}, x), \quad u_{n_0} < u''_{n_0} < u'_{n_0}, \\ &\leq \frac{-A''_k x}{\log x}, \end{aligned}$$

where  $A''_k$  depends only on  $k$ , since

$$F_k(u''_{n_0}, x) \leq A_k/u''_{n_0} \leq A_k/u_{n_0}$$

and  $(u'_{n_0} - u_{n_0})u_{n_0}^{-2}$  is the general term of a convergent series. Therefore,

$$0 < R \leq \frac{-A''_k x}{\log x} - \frac{A'_k x^{1/2}}{\log x}, \quad u'_{n_0} \leq x^{1/2} < u_{n_0} - 1.$$

We have shown that

$$Q + R = o(-1/\log x), \quad x \rightarrow 0,$$

and the construction is complete.

**24. Determining function the integral of a function of class  $L$ .** The theorem which we establish is little more than a corollary of Theorem 23.1.

**THEOREM 24.1.** *A necessary and sufficient condition that  $f(x)$  should have the representation*

$$(24.1) \quad f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} \phi(t) dt,$$

with  $\phi(t)$  of class  $L$  on  $(0, \infty)$ , is that  $f(x)$  should satisfy Conditions D and (23.2).

The conditions are necessary, by Theorems 19.1 and 23.1, since (24.1) can be written

$$\begin{aligned} f(x) &= \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\alpha(t), \\ (24.2) \quad \alpha(t) &= \int_{0+}^t \phi(u) du, \end{aligned}$$

$$(24.3) \quad \int_{0+}^{\infty} |d\alpha(t)| = \int_0^{\infty} |\phi(u)| du = M < \infty.$$

The conditions are sufficient. By Theorem 19.1

$$f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{\phi(u)du}{t+u}, \quad \int_0^{\infty} |\phi(u)| du = M < \infty;$$

this we may write as

$$f(x) = \int_{0+}^{\infty} \frac{dt}{x+t} \int_{0+}^{\infty} \frac{d\alpha(u)}{t+u},$$

where  $\alpha(t)$  is defined by (24.2) and has the property (24.3); and we have

$$(24.4) \quad \alpha(t) = \lim_{k \rightarrow \infty} \int_{0+}^t H_{k,u} [f(x)] du, \quad t > 0.$$

By Theorem 17.1

$$\int_0^{\infty} |H_{k,t} [f(x)]| dt \leq M, \quad k = 2, 3, \dots$$

Then  $f(x)$  satisfies the hypotheses of Theorem 23.1, and

$$(24.5) \quad f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} d\beta(t),$$

where

$$\beta(t) = \lim_{k \rightarrow \infty} \int_{0+}^t H_{k,u} [f(x)] du, \quad t > 0.$$

Comparing this with (24.4), we obtain

$$\beta(t) \equiv \alpha(t) = \int_0^t \phi(u) du,$$

and (24.5) reduces to (24.1).

**25. Determining function the integral of a bounded function.** We prove the following theorem:

**THEOREM 25.1.** *A necessary and sufficient condition that  $f(x)$  should have the representation*

$$(25.1) \quad f(x) = \int_{0+}^{\infty} \frac{\log(x/t)}{x-t} \phi(t) dt,$$

with  $\phi(t)$  bounded almost everywhere, is that  $f(x)$  should satisfy Conditions E, and that for an infinite sequence of integers  $k$ ,

$$(25.2) \quad \log x \left| \int_x^\infty t^{-1} H_{k,t} [f(x)] dt \right| \leq \epsilon(x), \quad x \geq 4,$$

where  $\lim_{x \rightarrow \infty} \epsilon(x) = 0$ , and  $\epsilon(x)$  is independent of  $k$ .

To establish the necessity of the conditions, we have only to establish (25.2), because of Theorems 20.1 and 5.2. We have

$$H_{k,t} [f(x)] = \int_0^\infty F_k(u, t) \phi(u) du, \quad k \geq 2.$$

Take  $x > 1$ . Then

$$\begin{aligned} \int_x^\infty t^{-1} H_{k,t} [f(x)] dt &= \int_x^\infty t^{-1} dt \left( \int_0^1 + \int_1^\infty \right) F_k(u, t) \phi(u) du \\ &= I_1 + I_2. \end{aligned}$$

To discuss  $I_2$ , set

$$\beta(u) = \int_u^\infty t^{-1} \phi(t) dt.$$

By Theorem 3.1,

$$(25.3) \quad \beta(u) = o(1/\log u), \quad u \rightarrow \infty;$$

then, since  $F_k(u, t) = O(1/u)$ , ( $u \rightarrow \infty$ ), we have

$$\begin{aligned} (25.4) \quad \int_1^\infty F_k(u, t) \phi(u) du &= - \int_1^\infty u F_k(u, t) d\beta(u) \\ &= \beta(1) F_k(1, t) + \int_1^\infty \beta(u) \frac{\partial}{\partial u} (u F_k(u, t)) du. \end{aligned}$$

By the homogeneity of  $F_k(u, t)$ ,

$$\frac{\partial}{\partial u} (u F_k(u, t)) = -t \frac{\partial}{\partial t} F_k(u, t);$$

therefore

$$(25.5) \quad \int_1^\infty \beta(u) \frac{\partial}{\partial u} (u F_k(u, t)) du = -t \int_1^\infty \beta(u) \frac{\partial}{\partial t} F_k(u, t) du.$$

It is easily verified that, for each  $k$ , the integral on the right is uniformly convergent for  $t \geq \delta > 0$ , and hence that we may take the symbol  $\partial/\partial t$  outside the integral sign. Using (25.5) in (25.4), then, we obtain

$$(25.6) \quad I_2 = \beta(1) \int_x^\infty t^{-1} F_k(1, t) dt - \int_1^\infty \beta(u) F_k(u, t) du \Big|_{t=x}^\infty = I_2' + I_2''.$$

Consider the integral

$$(25.7) \quad \int_1^{\infty} \beta(u) F_k(u, x) du = \left( \int_1^{x^{1/2}} + \int_{x^{1/2}}^{\infty} \right) \beta(u) F_k(u, x) du.$$

$$(25.8) \quad \left| \int_{x^{1/2}}^{\infty} \beta(u) F_k(u, x) du \right| \leq \underset{u \geq x^{1/2}}{\text{u.b.}} \left| \beta(u) \right| \int_{x^{1/2}}^{\infty} F_k(u, x) du \\ \leq \underset{u \geq x^{1/2}}{\text{u.b.}} \left| \beta(u) \right|.$$

There is a constant  $B$  such that  $|\beta(u)| \leq B$ , ( $1 \leq u < \infty$ ). Then

$$\begin{aligned} \left| \int_1^{x^{1/2}} \beta(u) F_k(u, x) du \right| &\leq B \int_1^{x^{1/2}} F_k(u, x) du \\ &= Bx \int_{x^{-1}}^{x^{-1/2}} F_k(vx, x) dv, & u = vx, \\ &= B \int_{x^{-1}}^{x^{-1/2}} F_k(v, 1) dv \\ &= B \int_{x^{-1}}^{x^{-1/2}} v F_k(1, v) dv \\ &\leq B \int_0^{x^{-1/2}} F_k(1, v) dv, & x > 1, \\ &= BH_k(x^{-1/2}). \end{aligned}$$

Applying Lemma 22.1, we then have

$$(25.9) \quad \left| \int_1^{x^{1/2}} \beta(u) F_k(u, x) du \right| \leq ABx^{-1/4}, \quad x \geq 4.$$

Now

$$\beta(u) = o(1/\log u), \quad u \rightarrow \infty.$$

Combining (25.9), (25.8), and (25.3) and referring to (25.7), (25.6), we see that for  $x \geq 4$

$$(25.10) \quad |I_2'| = \left| \int_1^{\infty} \beta(u) F_k(u, x) du \right| \leq \epsilon_1(x)/\log x,$$

where  $\epsilon_1(x) = o(1)$ , ( $x \rightarrow \infty$ ), and  $\epsilon_1(x)$  does not depend on  $k$ .

We have still to discuss

$$I_1 = \int_x^{\infty} t^{-1} dt \int_0^1 F_k(u, t) \phi(u) du$$

and

$$I_2' = \beta(1) \int_x^\infty t^{-1} F_k(1, t) dt.$$

For  $t > u$ ,  $F_k(u, t)$  is an increasing function of  $u$  (Lemma 7.4); therefore for  $t > 1$ ,

$$\left| \int_0^1 F_k(u, t) \phi(u) du \right| \leq M F_k(1, t),$$

where  $|\phi(u)| \leq M$  almost everywhere. Hence

$$|I_1| \leq M \int_x^\infty t^{-1} F_k(1, t) dt, \quad x > 1,$$

and

$$|I_1 + I_2'| \leq (M + |\beta(1)|) \int_x^\infty t^{-1} F_k(1, t) dt.$$

But for  $x > 1$ ,

$$\begin{aligned} \int_x^\infty t^{-1} F_k(1, t) dt &= \int_0^{x^{-1}} s^{-1} F_k(1, s^{-1}) ds, & st = 1, \\ &= \int_0^{x^{-1}} F_k(s, 1) ds = \int_0^{x^{-1}} s F_k(1, s) ds \\ &\leq \int_0^{x^{-1}} F_k(1, s) ds = H_k(x^{-1}). \end{aligned}$$

Again by Lemma 22.1, we see that

$$|I_1 + I_2'| \leq \epsilon_2(x)/\log x, \quad x \geq 4,$$

where  $\epsilon_2(x) = o(1)$ ,  $(x \rightarrow \infty)$ , and is independent of  $k$ . Combining this with (25.10), we have (25.2).

We now establish the sufficiency of our conditions. By Theorem 20.1, Conditions E imply that

$$(25.11) \quad f(x) = \int_{0+}^\infty \frac{dt}{x+t} \int_{0+}^\infty \frac{\phi(u) du}{t+u},$$

with  $|\phi(u)| \leq M$  almost everywhere. By Theorem 10.1

$$(25.12) \quad \phi(u) = \lim_{k \rightarrow \infty} H_{k,u}[f(x)]$$

for almost all  $u$ . Since (25.2) is satisfied, we have

$$\left| \int_x^\infty t^{-1} H_{k,t} [f(x)] dt \right| \leq \epsilon(x) / \log x, \quad x \geq 4,$$

for an infinite sequence of integers  $k$ , with  $\epsilon(x)$  independent of  $k$ , and  $\epsilon(x) = o(1)$ , ( $k \rightarrow \infty$ ). Then for  $k$  in the sequence,

$$\left| \int_{x_1}^{x_2} \frac{1}{t} H_{k,t} [f(x)] dt \right| \leq \frac{\epsilon(x_1)}{\log x_1} + \frac{\epsilon(x_2)}{\log x_2}, \quad x_2 \geq x_1 \geq 4.$$

Let  $k \rightarrow \infty$  in the sequence. By (25.12) and bounded convergence,

$$\left| \int_{x_1}^{x_2} \frac{\phi(t)}{t} dt \right| \leq \frac{\epsilon(x_1)}{\log x_1} + \frac{\epsilon(x_2)}{\log x_2}, \quad x_2 \geq x_1 \geq 4.$$

Therefore

$$\int_1^\infty t^{-1} \phi(t) dt$$

converges, and, if we set  $x_1 = x$  and let  $x_2 \rightarrow \infty$ , we obtain

$$(25.13) \quad \left| \int_x^\infty \frac{\phi(t)}{t} dt \right| \leq \frac{\epsilon(x)}{\log x} = o\left(\frac{1}{\log x}\right), \quad x \rightarrow \infty.$$

Moreover,

$$(25.14) \quad \left| \int_0^t \phi(u) du \right| \leq Mt = o(-1/\log t), \quad t \rightarrow 0.$$

Relations (25.13) and (25.14) are the conditions of Theorem 5.3; this theorem now permits us to change the order of integration in (25.11), obtaining the representation (25.1) for  $f(x)$ .

HARVARD UNIVERSITY,  
CAMBRIDGE, MASS.

# THE ABSTRACT GROUPS $G^{m,n,p*}$

BY  
H. S. M. COXETER

## CONTENTS

SECTION	PAGE
Preface.....	74
CHAPTER I. $(l, m   n, k)$	
1.1. Introduction; the polyhedral groups.....	77
1.2. The criterion for finiteness, when $l$ and $m$ are even.....	79
1.3. Particular cases.....	81
1.4. The criterion when $l$ and $k$ are even and $n = 2$ .....	82
1.5. Theorem A.....	82
1.6. Other groups that satisfy the criterion; Theorem B.....	83
1.7. Finite groups which violate the criterion.....	84
CHAPTER II. $(l, m, n; q)$	
2.1. Various forms of the abstract definition.....	86
2.2. The abelian case ( $q = 1$ ).....	87
2.3. The polyhedral groups and $(2, 3, 7; q)$ .....	88
2.4. A lemma on automorphisms.....	89
2.5. The derivation of $(2, m, m; q)$ and $(2, m, 2n; k)$ from $(4, m   2, 2q)$ and $(m, m   n, k)$ .....	90
2.6. The criterion for finiteness; Theorems C and D.....	93
2.7. The derivation of $(n, n, n; q)$ , $(3, 3, 3; k)$ , $(3, 3, 4; 3)$ from $(3, 3   n, 3q)$ , $(3, 3   3, k)$ , $(3, 3   4, 4)$ .....	95
2.8. Groups of genus one.....	97
2.9. An extension of the criterion; Fig. 1; Theorem D'.....	101
CHAPTER III. $G^{m,n,p}$	
3.1. The derivation of $G^{m,n,p}$ ( $p$ even) from its subgroup $(2, m, n; p/2)$ .....	104
3.2. The extended polyhedral groups.....	105
3.3. Cases in which $m, n$ are odd, while $p$ is even.....	107
3.4. Groups of genus one: $G^{4,4,2k}$ and $G^{3,6,2k}$ ; Fig. 2.....	109
3.5. The derivation of $G^{m,2n,2k}$ from its subgroup $(m, m   n, k)$ .....	111

\* Presented to the Society, September 10, 1937, under the title *A graphical representation for the simple group of order 504*; received by the editors September 8, 1937.

3.6. Definitions for $G^{m,n,p}$ ( $m$ odd) in terms of two generators.....	112
3.7. Cases in which $m, n, p$ are all odd.....	114
3.8. The criterion for finiteness; Theorems E and F; Figs. 3, 4.....	117
3.9. $G^{3,n,p}$ and the Fibonacci numbers; Theorem G; Fig. 5.....	121

#### CHAPTER IV. Graphical representation

4.1. Dyck's general group picture.....	124
4.2. The group picture for $G^{m,n,p}$ ; Figs. 6, 7.....	125
4.3. The regular map $\{m, n\}_p$ ; Fig. 8.....	127
4.4. The semi-regular map $\{m/n\}_p$ ; Fig. 9.....	128
4.5. The manner in which $\{m/n\}_p$ represents $G^{m,n,p}$ ; Figs. 10, 11....	130
4.6. Special cases; Figs. 12-22.....	131
4.7. The polyhedral groups.....	140
4.8. Polyhedra of higher connectivity; Fig. 23.....	140
4.9. The more general group $((b_1, b_2, \dots, b_r; p))$ .....	142
Table I. The known finite groups $(l, m   n, k)$ .....	146
Table II. The known finite groups $(l, m, n; q)$ .....	147
Table III. The known finite groups $G^{m,n,p}$ .....	147
Appendix. On the collapse of $(2, 3, 9; 4)$ and $(2, 5, 7; 2)$ .....	148
References.....	149

#### PREFACE

This paper treats of three families of abstract groups, defined by the following sets of relations:\*

$$(l, m | n, k): \quad R^l = S^m = (RS)^n = (R^{-1}S)^k = 1;$$

$$(l, m, n; q): \quad R^l = S^m = (RS)^n = (R^{-1}S^{-1}RS)^q = 1;$$

$$G^{m,n,p}: \quad A^m = B^n = C^p = (AB)^2 = (BC)^2 = (CA)^2 = (ABC)^2 = 1.$$

Here, " $R^l=1$ " means " $R$  is of period  $l$ ." When this and the remaining relations imply  $R^{1/d}=1$ , ( $d>1$ ), we acknowledge a case of *collapse: total* collapse if the relations reduce every generator to the identity, otherwise *partial*.

These groups are interrelated in various ways, of which the most important are as follows:  $(m, m | n, k)$  is a subgroup of index two in  $(2, m, 2n; k)$ ;  $(2, m, n; q)$  is a subgroup of index two in  $G^{m,n,2q}$ .

The study of these families is justified by the number of important groups which they contain. In particular, the families include the following *simple groups*:

\* The symbols are designed to admit the combination  $(l, m | n, k; q)$  (symmetrical between  $l$  and  $m$ , and between  $n$  and  $k$ ), which would mean the group defined by  $R^l = S^m = (RS)^n = (R^{-1}S)^k = (R^{-1}S^{-1}RS)^q = 1$ .

- $\mathfrak{P}_1(5)$  (order 60)  $\sim (5, 5 | 2, 3) \sim (2, 3 | 5, 5) \sim (2, 3, 5; 5) \sim G^{3,5,5}$ ,  
 $\mathfrak{P}_1(7)$  (order 168)  $\sim (3, 3 | 4, 4) \sim (4, 7 | 2, 3) \sim (2, 3, 7; 4)$ ,  
 $\mathfrak{P}_1(9)$  (order 360)  $\sim (5, 5 | 2, 4) \sim (3, 4, 5; 2)$ ,  
 $\mathfrak{P}_1(8)$  (order 504)  $\sim G^{3,7,9}$ ,  
 $\mathfrak{P}_1(11)$  (order 660)  $\sim G^{5,5,5}$ ,  
 $\mathfrak{P}_1(13)$  (order 1092)  $\sim (l, 7 | 2, 3) \sim (2, 3, 7; l)$ , ( $l=6$  or  $7$ ),  
 $\mathfrak{P}_1(17)$  (order 2448)  $\sim (4, 9 | 2, 3)$ ,  
 $\mathfrak{P}_1(19)$  (order 3420)  $\sim (2, 5, 9; 2) \sim G^{3,9,9}$ ,  
 $\mathfrak{P}_1(23)$  (order 6072)  $\sim (2, 3, 11; 4)$ ,  
 $\mathfrak{P}_1(29)$  (order 12180)  $\sim G^{3,7,15}$ .

Following Schreier and van der Waerden,\* I let  $\mathfrak{P}_1(q)$  denote the group commonly called  $LF(2, q)$  or  $PSL(2, q)$ ,  $q$  being a prime or a prime power. When  $q$  is odd, I let  $\tilde{\mathfrak{P}}_1(q)$  denote the group of all linear fractional transformations in the Galois field of order  $q$ , otherwise called  $PGL(2, q)$ ; this contains  $\mathfrak{P}_1(q)$  as a subgroup of index two. The groups  $\mathfrak{P}_1(9)$  and  $\tilde{\mathfrak{P}}_1(9)$ , being the alternating and symmetric groups of degree six, will usually be denoted by  $G_{61/2}$  and  $G_{61}$ .

The above definitions for  $\mathfrak{P}_1(5)$  (or  $G_{51/2}$ ) are merely redundant forms of Hamilton's definition†

$$\iota^2 = \kappa^3 = (\iota\kappa)^5 = 1.$$

The first definition for  $\mathfrak{P}_1(7)$  is due to G. A. Miller; the second is immediately deducible from Burnside's; the third is essentially Dyck's‡ (but was put into precisely this form by H. R. Brahana§). The second definition for  $\mathfrak{P}_1(13)$  is due to Brahana§ ( $l=6$ ) and Sinkov|| ( $l=7$ ); the first is easily deduced from it. The first definition for  $\mathfrak{P}_1(9)$  (or  $G_{61/2}$ ), and the definitions for  $\mathfrak{P}_1(8)$  and  $\mathfrak{P}_1(17)$ , have already been published;¶ but  $(3, 4, 5; 2)$ ,  $G^{5,5,5}$ ,  $(2, 5, 9; 2)$ ,  $G^{3,9,9}$ ,  $(2, 3, 11; 4)$ ,  $G^{3,7,15}$  are quite new.

We shall see (Theorem G) that, for every prime  $p$ , the linear fractional group  $\mathfrak{P}_1(p)$  is a factor group either of some  $(2, 3, p; q)$  or of some  $G^{3,n,p}$ . This result may be compared with the following.\*\* For the group  $\mathfrak{P}_1(2^m)$ , every pair of generators of periods two and three satisfies the definition of some  $G^{3,n,p}$ .

\* Schreier and van der Waerden [1]. Their Theorem 1 shows that  $\mathfrak{P}_1(p)$  is the group of isomorphisms of  $\mathfrak{P}_1(p)$  ( $p$  prime). We shall make frequent use of this theorem.

† Hamilton [1].

‡ Miller [1], p. 364; Burnside [1], p. 422; Dyck [1], p. 41.

§ Brahana [2], pp. 351, 354.

|| Sinkov [1], p. 239.

¶ Todd and Coxeter [1], p. 31; Sinkov [3], p. 70; Coxeter [7], p. 56.

\*\* Sinkov [6], p. 454.

The following seven theorems will be proved:

**THEOREM A.** *The groups  $(2, m|n, k)$  for  $n \neq k$ ,  $(3, m|2, k)$  for  $m \neq k$ ,  $(5, m|2, 3)$  for  $m \neq 5$ , and  $(l, m|2, 2)$  for  $l > 2$  and  $m$  odd, all collapse. Apart from these cases, if  $l$  and  $m$  are even, or if  $l$  and  $k$  are even and  $n=2$ , the group  $(l, m|n, k)$  is finite when  $2 \sin \pi/l \sin \pi/m > \cos \pi/n + \cos \pi/k$ , and infinite otherwise.*

**THEOREM B.** *For all infinite groups  $(l, m|n, k)$ ,*

$$2 \sin \pi/l \sin \pi/m \leq \cos \pi/n + \cos \pi/k.$$

**THEOREM C.** *If  $q > 1$  and  $1/m + 1/n \leq 1/2$ , and  $m$  and  $n$  are either even or equal (or both), the group  $(2, m, n; q)$  is finite when*

$$\cos 2\pi/m + \cos 2\pi/n + \cos \pi/q < 1,$$

*and infinite otherwise.*

**THEOREM D.** *For all infinite groups  $(2, m, n; q)$ ,*

$$\cos 2\pi/m + \cos 2\pi/n + \cos \pi/q \geq 1.$$

**THEOREM E.** *If the smallest of  $m, n, p$  is greater than 2, while the next is greater than 3, and if these three numbers are either all even, or one even and the other two equal, the group  $G^{m,n,p}$  is finite when*

$$\cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p < 1,$$

*and infinite otherwise.*

**THEOREM F.** *For all (noncollapsing) groups  $G^{m,n,p}$ , save  $G^{2,n,2n}$  ( $n$  odd),*

$$(1 - \cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p)(1 + \cos 2\pi/m - \cos 2\pi/n + \cos 2\pi/p)(1 + \cos 2\pi/m + \cos 2\pi/n - \cos 2\pi/p) \geq 0;$$

*and for all infinite groups  $G^{m,n,p}$ ,*

$$\cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p \geq 1.$$

**THEOREM G.** *If  $p$  is a prime, congruent to 1 or 3 (mod 4), the group  $\mathbb{P}_1(p)$  or  $\mathbb{P}_1(p)$  (respectively) is a factor group of  $G^{3,n,p}$ , where  $n$  is the ordinal of the first Fibonacci number that is divisible by  $p$ . When  $p \equiv 3 \pmod{4}$ ,  $\mathbb{P}_1(p)$  is a factor group of  $(2, 3, p; n/2)$ .*

In the final section, we see that  $G^{m,n,p}$  has an elegant representation on a "semi-regular map" of  $m$ -gons and  $n$ -gons (or  $n$ -gons and  $p$ -gons, or  $p$ -gons and  $m$ -gons). Although the representation has not (so far) given any new information about the groups, it provides an alternative method for enumerating the operators, whenever the order does not greatly exceed five hundred.

Moreover, the representation seems to clarify the phenomenon of "collapse."

Tables I, II, III, at the end of the paper, summarize the special results, and at the same time serve as an index. In Table III, it is noteworthy that the group  $G^{3,7,p}$ , with  $p=12$  or  $13$  or  $14$ , has order  $12 \times 13 \times 14$ . (Although  $G^{3,7,12} \sim G^{3,7,14}$ ,  $G^{3,7,13}$  is a different group of the same order.)

A considerable part of this work is the fruit of discussions with Dr. A. Sinkov, extending over several years. To him I would express my sincere gratitude.

#### CHAPTER I. $(l, m | n, k)$

**1.1. Introduction; the polyhedral groups.** The group  $(l, m | n, k)$ , which is defined by

$$(1.11) \quad R^l = S^m = (RS)^n = (R^{-1}S)^k = 1,$$

may be regarded as a factor group of

$$(1.111) \quad R^l = S^m = T^n = RST = 1.$$

In terms of  $S$  and  $T$ , it takes the form

$$(1.12) \quad S^m = T^n = (ST)^l = (S^2T)^k = 1.$$

The group (1.111) is known\* to be finite when

$$\frac{1}{l} + \frac{1}{m} + \frac{1}{n} > 1,$$

and infinite otherwise. When  $n=2$ , we denote it by  $[l, m]'$ , since it is the rotation group of the regular polyhedron†  $\{l, m\}$ . Its factor group  $(l, m | 2, k)$  is the rotation group of the regular skew polyhedron‡  $\{l, m | k\}$ .

The dihedral group  $[2, m]'$ , defined by  $R^2 = T^2 = (RT)^m = 1$ , may be denoted more briefly by  $[m]$ , since it is the complete group of the regular  $m$ -gon  $\{m\}$ . Finally,  $[m]'$  denotes the cycle group of order  $m$ , defined by  $S^m = 1$ .

By interchanging  $R$  and  $S$  in (1.11), we get  $(l, m | n, k) \sim (m, l | n, k)$ . Again, writing  $R^{-1}$  for  $R$ , we have  $(l, m | n, k) \sim (l, m | k, n)$ . Hence

$$(1.13) \quad (l, m | n, k) \sim (m, l | n, k) \sim (m, l | k, n) \sim (l, m | k, n).$$

When naming special groups, we shall usually arrange the symbols so that  $l \leq m, n \leq k$ .

\* Threlfall [1], p. 28 ( $r=3$ ).

† Todd [1], p. 214. When  $1/l + 1/m < 1/2$ , the polyhedron can be constructed in the "Minkowskian" space whose dimensions are two space-like and one time-like; see Coxeter [4], p. 24.

‡ Coxeter [7], p. 48.

When  $l=2$ , (1.11) takes the form

$$R^2 = S^n = (RS)^n = (RS)^k = 1.$$

Hence

$$(1.14) \quad (2, m \mid n, k) \text{ collapses (partially or totally) if } n \neq k,$$

but

$$(1.141) \quad (2, m \mid n, n) \sim [m, n]'$$

(finite when  $1/m + 1/n > 1/2$ ). In particular,

$$(1.142) \quad (2, 2 \mid n, n) \sim (2, n \mid 2, 2) \sim [n]$$

(where  $[n]$  is the dihedral group of order  $2n$ ).

When  $l=3$ , (1.11) takes the form

$$R^3 = S^m = (RS)^n = (R^{-1}S)^k = 1,$$

where, of course,  $R^{-1}$  may be replaced by  $R^2$ . Putting  $S' = RS$ , we derive

$$R^3 = (R^{-1}S')^m = S'^n = (RS')^k = 1.$$

Hence

$$(1.15) \quad (3, m \mid n, k) \sim (3, n \mid k, m) \sim (3, k \mid m, n).$$

By (1.14) and (1.15),

$$(1.16) \quad (3, m \mid 2, k) \text{ collapses if } m \neq k,$$

but

$$(1.161) \quad (3, m \mid 2, m) \sim [3, m]'$$

(finite when  $m < 6$ ).

When  $m=5$  and  $n=2$ , (1.12) takes the form

$$S^5 = T^2 = (ST)^l = (S^2T)^k = 1.$$

Putting  $S' = S^2$ , we derive

$$S'^5 = T^2 = (S'^3T)^l = (S'T)^k = 1.$$

Here,  $S'^3T$  may be replaced by  $S'^2T$  (conjugate to its inverse). Hence

$$(1.17) \quad (l, 5 \mid 2, k) \sim (k, 5 \mid 2, l).$$

By (1.16) and (1.17),

$$(1.18) \quad (k, 5 \mid 2, 3) \text{ collapses if } k \neq 5,$$

but

$$(1.181) \quad (5, 5 | 2, 3) \sim [3, 5]'$$

(where  $[3, 5]'$  is the icosahedral group).

Consider now the group  $(l, m | 2, 2)$ , defined by

$$R^l = S^m = (RS)^2 = (R^{-1}S)^2 = 1.$$

Let  $U = SR$ ,  $V = R^{-1}S$ , and suppose that  $m = 2q + 1$ . Then

$$S^{-1} = S^{m-1} = S^{2q} = (UV)^q,$$

and

$$R = S^{-1}U = (UV)^q U.$$

Since  $U^2 = V^2 = 1$ , it follows that  $R^2 = 1$ . Hence

$$(1.19) \quad (l, m | 2, 2) \text{ collapses if } l > 2, m \text{ odd.}$$

We have already seen (1.142) that  $(2, m | 2, 2) \sim [m]$ . The case when  $l$  and  $m$  are both even will be considered for general values of  $n$  and  $k$ .

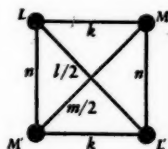
**1.2. The criterion for finiteness, when  $l$  and  $m$  are even.** The group  $(l, m | n, k)$  clearly possesses an automorphism which replaces the generators  $R$  and  $S$  by their inverses. Let  $T$  denote an involutory operator which transforms the group according to this automorphism. By adjoining  $T$ , we derive the new group

$$(1.21) \quad R^l = S^m = T^2 = (RT)^2 = (ST)^2 = (RS)^n = (R^{-1}S)^k = 1,$$

whose order is twice that of  $(l, m | n, k)$ .

Now suppose that  $l$  and  $m$  are even, and consider the group generated by reflections\*

$$(1.22) \quad \begin{aligned} L^2 = L'^2 = M^2 = M'^2 &= (LL')^{l/2} = (MM')^{m/2} \\ &= (LM')^n = (ML')^n = (LM)^k = (L'M')^k = 1. \end{aligned}$$



This possesses an automorphism which interchanges  $L$  and  $L'$ ,  $M$  and  $M'$ . Let  $T$  denote an involutory operator which transforms the group according to this automorphism. We adjoin  $T$  by inserting the extra relations

\* Coxeter [4]. In that paper, Lemmas 1 and 2 are unnecessary. In the proof of Lemma 3, after obtaining the invariant form  $\sum c_{ij}y_i^2$ , we can introduce a euclidean metric by writing  $x_j = c_j^{1/2}y_j$ .

$$T^2 = 1, \quad LT = TL' = R \text{ (say),} \\ MT = TM' = S \text{ (say).}$$

The generators  $L, L', M, M'$  may now be eliminated by substituting

$$L = RT, \quad L' = TR, \quad M = ST, \quad M' = TS;$$

and the augmented group takes the form

$$T^2 = (RT)^2 = (ST)^2 = R^l = S^m \\ = (RS)^n = (R^{-1}S)^k = 1,$$

which is the same as (1.21).

If  $l$  and  $m$  are even,  $(l, m | n, k)$  has the same order as the above group generated by reflections, both being subgroups of index two in the same larger group.

This result enables us to assert that  $(l, m | n, k)$  never collapses, as long as  $l$  and  $m$  are even and greater than 2 (see (1.14)\*). Moreover, it provides a definite criterion for finiteness. For we know† that the group generated by reflections (1.22) is finite or infinite according as  $\Delta$  is or is not positive, where

$$\Delta = \begin{vmatrix} 1 & -\cos 2\pi/l & -\cos \pi/k & -\cos \pi/n \\ -\cos 2\pi/l & 1 & -\cos \pi/n & -\cos \pi/k \\ -\cos \pi/k & -\cos \pi/n & 1 & -\cos 2\pi/m \\ -\cos \pi/n & -\cos \pi/k & -\cos 2\pi/m & 1 \end{vmatrix} \\ = \{ (1 + \cos 2\pi/l)(1 + \cos 2\pi/m) - (\cos \pi/n - \cos \pi/k)^2 \} \\ \cdot \{ (1 - \cos 2\pi/l)(1 - \cos 2\pi/m) - (\cos \pi/n + \cos \pi/k)^2 \} \\ = (2 \cos \pi/l \cos \pi/m + |\cos \pi/n - \cos \pi/k|) \\ \cdot (2 \cos \pi/l \cos \pi/m - |\cos \pi/n - \cos \pi/k|) \\ \cdot (2 \sin \pi/l \sin \pi/m + \cos \pi/n + \cos \pi/k) \\ \cdot (2 \sin \pi/l \sin \pi/m - \cos \pi/n - \cos \pi/k).$$

Of these four factors, the first three are essentially positive ( $l, m > 2$ ), the first and third obviously, and the second because  $l$  and  $m$ , being even, must be at least 4, so that

$$2 \cos \pi/l \cos \pi/m \geq 1 \geq |\cos \pi/n - \cos \pi/k|.$$

Thus the criterion for finiteness reduces to

$$(1.23) \quad 2 \sin \pi/l \sin \pi/m > \cos \pi/n + \cos \pi/k.$$

\* If  $l=2$ , the g.g.r. has only three generators, since  $L'=L$ . It follows, even from the present point of view, that, to avoid collapse, we must have  $n=k$ .

† Coxeter [1], p. 137; [2], p. 602; [4], p. 24.

1.3. **Particular cases.** The finite groups, according to this criterion, are

$$(2i, 2j | 2, 2), (4, 4 | 2, k), (4, 6 | 2, 3), (4, 8 | 2, 3).$$

The orders of these are equal to the orders of the groups generated by reflections\*

$$[i, 2, j], [k, 2, k], [3, 3, 3], [3, 4, 3],$$

namely  $4ij, 4k^2, 120, 1152$ .

The group  $(2i, 2j | 2, 2)$ , defined by

$$S^{2i} = T^2 = (ST)^{2i} = (S^2T)^2 = 1,$$

has an invariant cyclic subgroup of order  $j$  (generated by  $S^2$ ), whose quotient group is the dihedral group of order  $4i$ . (It has also, of course, a cyclic subgroup of order  $i$ , whose quotient group is the dihedral group of order  $4j$ .) Returning to the definition in terms of  $R$  and  $S$ , we easily verify that  $(2i, 2j | 2, 2)$  is generated by the permutations†

$$(1.31) \quad \begin{cases} R = (a_1 a_2 \cdots a_{2i})(b_1 b_{2i})(b_2 b_{2i-1}) \cdots (b_i b_{i+1}), \\ S = (b_1 b_2 \cdots b_{2j})(a_1 a_{2i})(a_2 a_{2i-1}) \cdots (a_i a_{i+1}). \end{cases}$$

The group  $(4, 4 | 2, k)$  is well known as one of the groups of genus one.‡ It is generated (when  $k > 2$ ) by the permutations

$$(1.32) \quad R = (1 \ 2 \ 3 \ 4)(5 \ 6 \ 7 \ 8) \cdots, \quad S = (1 \ 2)(3 \ 4 \ 5 \ 6) \cdots,$$

where the final cycles (of two or four numbers) end with  $2k$ .

The group  $(4, 6 | 2, 3)$  is the symmetric group of degree five,§ generated by the permutations  $(1 \ 4 \ 2 \ 5), (1 \ 5)(2 \ 3 \ 4)$ .

The group  $(4, 8 | 2, 3)$  has a central of order two|| whose quotient group is generated by the permutations  $(1 \ 8)(2 \ 7 \ 3 \ 6)(4 \ 5), (1 \ 6 \ 3 \ 7 \ 4 \ 5 \ 2 \ 8)$ .

For all larger values of  $n$  and  $k$ , and for all larger even values of  $l$  and  $m$ ,  $(l, m | n, k)$  is infinite; for example,  $(4, 4 | 3, 3), (4, 6 | 2, 4), (6, 6 | 2, 3)$ , and  $(4, 10 | 2, 3)$  are infinite.

\* Todd [1], p. 224, (4). Their graphical symbols (Coxeter [4], p. 21) being disconnected, the first two of these are direct products of dihedral groups, namely  $[i] \times [j], [k] \times [k]$ .

† In this case

$$L = (a_1 a_{2i-1})(a_2 a_{2i-2}) \cdots (a_{i-1} a_{i+1}), \quad L' = (a_2 a_{2i})(a_3 a_{2i-1}) \cdots (a_i a_{i+2}),$$

$M$  and  $M'$  are analogous, with  $b$  and  $j$  in place of  $a$  and  $i$ , and

$$T = (a_1 a_{2i})(a_2 a_{2i-1}) \cdots (a_i a_{i+1})(b_1 b_{2j})(b_2 b_{2j-1}) \cdots (b_j b_{j+1}).$$

‡ Edington [1], pp. 197-202. In this case  $L = (2 \ 4)(6 \ 8) \cdots, M = (4 \ 6)(8 \ 10) \cdots, L' = (1 \ 3)(5 \ 7) \cdots, M' = (3 \ 5)(7 \ 9) \cdots$ , and  $T = (1 \ 2)(3 \ 4)(5 \ 6) \cdots$ .

§ In this case  $L = (1 \ 2), M = (2 \ 3), M' = (3 \ 4), L' = (4 \ 5)$ , and  $T = (1 \ 5)(2 \ 4)$ . For the geometrical aspect, see Coxeter [7], p. 49.

|| Since the polytope  $\{3, 4, 3\}$  has central symmetry.

**1.4. The criterion when  $l$  and  $k$  are even and  $n=2$ .** We observe that the criterion (1.23) remains valid when  $l=2$  and  $n=k$ ; it then has wider scope, since  $m$  may be odd. Moreover, it agrees with the fact that  $(3, m|2, m)$  is finite only when  $m < 6$ . This suggests its possible significance in yet other cases. We shall see that it is applicable to the whole class of groups  $(l, m|2, k)$  for which  $l$  and  $k$  are even, and greater than 2, while  $m > 3$  (but may be odd).

According to the criterion,  $(l, m|2, k)$  should be finite when  $l=k=4$  and  $m=5$ , but infinite for all greater even values of  $l$  and  $k$ , and for all greater values of  $m$ . Elsewhere\* we saw that the group  $(2p_1, m|2, 2p_2)$  is finite when  $m=5$  and  $p_1=p_2=2$ , but infinite for all greater values of  $m, p_1, p_2$ . Since the argument is rather subtle, an outline of it is repeated here.

From the group generated by reflections

$$(1.41) \quad R_i^2 = (R_i R_{i+1})^{l/2} = (R_i R_{i+2})^{k/2} = 1, \quad R_{i+m} = R_i$$

(implying  $l=k$  when  $m=3$ ), we derive the group

$$S^m = R_0^2 = (R_0 S^{-1} R_0 S)^{l/2} = (R_0 S^{-2} R_0 S^2)^{k/2} = 1,$$

whose order is  $m$  times as great, by adjoining an operator  $S$ , of period  $m$ , which cyclically permutes the  $R$ 's (so that  $R_i = S^{-i} R_0 S^i$ ). The augmented group and  $(l, m|2, k)$  are of the same order, having a common subgroup of index two. The group (1.41), and so also  $(l, m|2, k)$ , is infinite when  $m > 5$ , since  $R_i R_{i+3}$  is then of unspecified (and therefore infinite) period. It is known also to be infinite when  $m=5$  and  $l/2$  or  $k/2 > 2$ .

When  $m=5$  and  $l/2=k/2=2$ , (1.41) is the abelian group of order 32 and type  $(1, 1, 1, 1, 1)$ , generated (say) by the transpositions

$$R_i = (i \ i+5), \quad i = 0, 1, 2, 3, 4.$$

Hence†  $(4, 5|2, 4)$ , of order  $5 \times 32 = 160$ , is generated (in the form (1.12)) by the permutations

$$(1.42) \quad S = (0 \ 1 \ 2 \ 3 \ 4)(5 \ 6 \ 7 \ 8 \ 9), \quad T = (1 \ 9)(2 \ 8)(3 \ 7)(4 \ 6).$$

(The "common subgroup" is generated by  $S$  and  $(0 \ 9 \ 8 \ 7 \ 6)(5 \ 4 \ 3 \ 2 \ 1)$ ;  $T$  interchanges these.)

**1.5. Theorem A.** We summarize the results of §1.1, §1.2, and §1.4 in the following theorem:

**THEOREM A.** *The groups  $(2, m|n, k)$  for  $n \neq k$ ,  $(3, m|2, k)$  for  $m \neq k$ ,  $(5, m|2, 3)$  for  $m \neq 5$ , and  $(l, m|2, 2)$  for  $l > 2$  and  $m$  odd, all collapse. Apart*

\* Coxeter [6], p. 284, (4.9).

† Coxeter [6], p. 284.

from these cases, if  $l$  and  $m$  are even, or if  $l$  and  $k$  are even\* and  $n=2$ , the group  $(l, m | n, k)$  is finite when

$$2 \sin \pi/l \sin \pi/m > \cos \pi/n + \cos \pi/k,$$

and infinite otherwise.

The groups which this theorem shows to be finite are

$$\begin{aligned} &(l, m | 2, 2) \text{ (} l \text{ and } m \text{ even), } (2, 2 | n, n), \\ &(2, 4 | 3, 3), \quad (3, 4 | 2, 4), \quad (4, 4 | 2, k), \\ &(4, 5 | 2, 4), \quad (4, 6 | 2, 3), \quad (4, 8 | 2, 3). \end{aligned}$$

(See Table I, at the end of the paper.)

1.6. **Other groups that satisfy the criterion.** The remaining solutions of (1.23) are

$$\begin{aligned} &(2, 3 | 4, 4), \quad (2, 3 | 5, 5), \quad (2, 5 | 3, 3), \\ &(3, m | 3, 3) \text{ (} m < 6), \quad (3, 3 | 3, k), \\ &(3, 3 | 4, 4), \quad (3, 4 | 3, 4), \quad (3, 5 | 2, 5), \\ &(4, 5 | 2, 5), \quad (4, 7 | 2, 3), \quad (5, 5 | 2, 3). \end{aligned}$$

By (1.141),

$$(1.61) \quad (2, 3 | 4, 4) \sim [3, 4]' \sim G_{48},$$

the octahedral group. Similarly, by (1.141), (1.161), and (1.181),

$$(1.62) \quad (2, 5 | 3, 3) \sim (2, 3 | 5, 5) \sim (3, 5 | 2, 5) \sim (5, 5 | 2, 3) \sim G_{51/2},$$

the icosahedral group.

The group  $(3, 3 | 3, k)$  is another of the groups of genus one.† It is generated by the permutations

$$(1.63) \quad (1 \ 2 \ 3)(4 \ 5 \ 6) \cdots (\cdots 3k), \quad (3k \ 1 \ 2)(3 \ 4 \ 5) \cdots (\cdots 3k - 1),$$

and is of order  $3k^2$ . By (1.15),

$$(1.64) \quad (3, k | 3, 3) \sim (3, 3 | 3, k).$$

Miller‡ proved that  $(3, 3 | 4, 4)$  is Klein's 168-group, generated by permutations of the form

$$(1.65) \quad R = (0 \ 1 \ \infty)(2 \ 6 \ 4), \quad S = (0 \ 2 \ 3)(1 \ 6 \ 5),$$

\* Perhaps  $(5, m | 2, 3)$  should not have been mentioned, since its  $l$  and  $k$  are odd; but it seemed desirable to bring together all the known cases of collapse.

† Edington [1], p. 208.

‡ Miller [1], p. 364.

which are equivalent to the linear fractional substitutions

$$\left(\frac{1}{1-x}\right), \quad (4x+2) \pmod{7}.$$

By enumerating cosets\* of the octahedral subgroup generated by  $RSR$  and  $S$ , we easily verify the order 168 and obtain the alternative representation

$$R = (1\ 2\ 3)(4\ 5\ 6), \quad S = (2\ 3\ 4)(5\ 7\ 6).$$

The group  $(4, 7 | 2, 3)$  is the same, since it can be derived from Burnside's†

$$S_2^2 = S_7^7 = (S_7 S_2)^3 = (S_7^4 S_2)^4 = 1$$

by putting  $S_7 = S^2$ ,  $S_2 = T$ . Thus

$$(1.66) \quad (3, 4 | 3, 4) \sim (3, 3 | 4, 4) \sim (4, 7 | 2, 3) \sim \mathfrak{P}_1(7).$$

Finally, by (1.17),

$$(1.67) \quad (4, 5 | 2, 5) \sim (5, 5 | 2, 4) \sim G_{61/2},$$

$(5, 5 | 2, 4)$  having been used‡ as an example to illustrate the "Todd-Coxeter method" for enumerating cosets.

These results show that  $(l, m | n, k)$  is finite (or collapses) whenever (1.23) is satisfied; that is, whenever

$$2 \sin \pi/l \sin \pi/m > \cos \pi/n + \cos \pi/k.$$

Hence we obtain the following theorem:

**THEOREM B.** *For all infinite groups  $(l, m | n, k)$ ,*

$$2 \sin \pi/l \sin \pi/m \leq \cos \pi/n + \cos \pi/k.$$

**1.7. Finite groups which violate the criterion.** The converse is false, since  $(5, 5 | 2, 4)$ , which violates (1.23), is the same group as  $(4, 5 | 2, 5)$ , which satisfies it. So too,  $(3, m | 3, 3)$  violates (1.23) when  $m \geq 6$ , although  $(3, 3 | 3, k)$  satisfies it universally. We proceed to describe six further cases, namely:

$$(1.71) \quad (3, 3 | 4, 5) \sim (3, 4 | 3, 5) \sim (3, 5 | 3, 4), \text{ of order } 1080,$$

$$(1.72) \quad (6, 7 | 2, 3) \sim (7, 7 | 2, 3) \sim \mathfrak{P}_1(13), \S$$

$$(1.73) \quad (4, 9 | 2, 3) \sim \mathfrak{P}_1(17).$$

\* Todd and Coxeter [1].

† Burnside [1], p. 422.

‡ Todd and Coxeter [1], p. 31, (5).

§ The fact that  $(6, 7 | 2, 3)$  and  $(7, 7 | 2, 3)$  are finite, while  $(6, 6 | 2, 3)$  is infinite, shows that any attempt to "improve" the criterion would be futile.

The group  $(3, 3|4, 5)$  shows the efficiency of the Todd-Coxeter method so clearly that I shall perhaps be forgiven for writing out the work in full. Taking the icosahedral subgroup generated by  $RSR$  and  $S$  (and using rows instead of columns), we proceed with the enumeration of cosets as follows:

$R$	$R$	$R$	$S$	$S$	$S$	$R$	$S$	$R$	$S$	$R$	$S$	$R$	$S$
1	2	3	1	1	1	1	2	3	1	1	2	3	1
4	5	6	4	2	3	4	5	7	8	6	4	2	3
7	8	9	7	5	7	10	5	11	12	13	14	12	10
10	11	12	10	6	11	8	6	16	17	9	7	10	11
13	14	15	13	9	16	17	9	15	13	14	15	15	13
16	17	18	16	12	13	14	12	18	16	17	18	18	16
				15	15	15	15						
				18	18	18	18						

$R^{-1}$	$S$	$R^{-1}$	$S$	$R^{-1}$	$S$	$R^{-1}$	$S$	$R^{-1}$	$S$
1	3	4	6	11	10	5	4	2	1
3	2	3	2	3	2	3	2	3	2
7	9	16	18	18	17	9	8	6	5
10	12	13	15	15	14	12	11	8	7
14	13	14	13	14	13	14	13	14	13
17	16	17	16	17	16	17	16	17	16

This verifies the order,  $18 \times 60 = 1080$ , which was first obtained by Miller,\* and gives us the permutations

$$(1.74) \quad \begin{cases} R = (1\ 2\ 3)(4\ 5\ 6)(7\ 8\ 9)(10\ 11\ 12)(13\ 14\ 15)(16\ 17\ 18), \\ S = (2\ 3\ 4)(5\ 7\ 10)(6\ 11\ 8)(9\ 16\ 17)(12\ 13\ 14). \end{cases}$$

(To see how much labor has been saved, compare pages 365–368 of Miller's paper.) The permutation

$$(1.741) \quad \begin{aligned} Z &= (1\ 15\ 18)(2\ 13\ 16)(3\ 14\ 17)(4\ 12\ 9)(5\ 10\ 7)(6\ 11\ 8) \\ &= (R^{-1}S^{-1}RS)^5 \end{aligned}$$

generates the central, of order three, whose quotient group† is the alternating group  $G_{61/2}$ , generated by  $(1\ 2\ 3)(4\ 5\ 6)$  and  $(2\ 3\ 4)$ .

The groups  $(6, 7|2, 3)$  and  $(7, 7|2, 3)$  are equivalent to definitions for  $\mathfrak{P}_1(13)$  which were given by Brahana and Sinkov, respectively.‡ Identifying

\* Miller [1], p. 368.

†  $(3, 3|4, 5; 5)$ , in the notation of the footnote to p. 74 above.

‡ Brahana [2], p. 354; Sinkov [1], p. 239; Coxeter [7], p. 56.

our  $T$  with theirs, namely

$$(1.75) \quad T = (1\ 12)(2\ 11)(3\ 10)(4\ 9)(5\ 8)(6\ 7),$$

we find that the other generator of  $(6, 7|2, 3)$  or  $(7, 7|2, 3)$  (namely,  $S$ ) is, in Sinkov's notation,\*  $(S_{11}T)^3$  or  $(S_{13}T)^3$ , respectively. Thus  $(6, 7|2, 3)$  is generated by  $T$  and

$$(1.751) \quad S = (0\ 3\ 10\ 8\ 2\ 5\ 9)(1\ 6\ 7\ \infty\ 11\ 12\ 4),$$

while  $(7, 7|2, 3)$  is generated by the same  $T$  and

$$(1.752) \quad S = (0\ 12\ 9\ 1\ 11\ 10\ 5)(2\ 3\ 7\ 8\ 6\ \infty\ 4).$$

In the case of  $(4, 9|2, 3)$ , by enumerating the 272 cosets of the cyclic subgroup generated by  $S$ , I found the order to be 2448. This irresistibly suggested (1.73), which Sinkov showed to be true.† He cited two linear fractional substitutions (mod 17) which lead to the permutations

$$S = (0\ 6\ 10\ 14\ 3\ 4\ 15\ 5\ 16)(1\ 13\ \infty\ 7\ 2\ 12\ 11\ 9\ 8),$$

$$T = (1\ 16)(2\ 15)(3\ 14)(4\ 13)(5\ 12)(6\ 11)(7\ 10)(8\ 9).$$

Although there are, as we have seen, infinitely many finite groups which violate the criterion (1.23), it still seems improbable that all the unknown groups  $(l, m|n, k)$  will turn out to be finite. On the contrary, my conjecture is that they are all infinite. We shall see later (2.67) that  $(3, 3|4, 6)$  is certainly infinite.

## CHAPTER II. $(l, m, n; q)$

**2.1. Various forms of the abstract definition.** The group  $(l, m, n; q)$ , defined by

$$(2.11) \quad R^l = S^m = (RS)^n = (R^{-1}S^{-1}RS)^q = 1,$$

involves the numbers  $l, m, n$  symmetrically, since its definition is equivalent to

$$(2.111) \quad R^l = S^m = T^n = RST = (TSR)^q = 1.$$

When  $n=2$ , so that  $RS=S^{-1}R^{-1}$ , we have‡  $RS^{-1}R^{-1}S=R^2S^2$ . Hence  $(2, l, m; q)$  is defined by

$$(2.12) \quad R^l = S^m = (RS)^2 = (R^2S^2)^q = 1.$$

When  $m=3$  and  $n=2$ , so that  $TR^{-1}T=RTR$ , we have  $TR^{-1}TR=RTR^2$ ,

\* Sinkov [3], p. 71. I have replaced Sinkov's symbols 13 and 14 by 0 and  $\infty$ , respectively.

† Coxeter [7], p. 56.

‡ Edington [1], p. 195.

which is conjugate to  $R^3T$ . Thus the definition

$$R^l = T^2 = (TR)^3 = (TR^{-1}TR)^q = 1$$

for  $(2, 3, l; q)$  is equivalent to

$$(2.13) \quad R^l = T^2 = (RT)^3 = (R^3T)^q = 1.$$

**2.2. The abelian case ( $q=1$ ).** We proceed to prove that  $(l, m, n; 1)$  collapses unless it is of the form  $(bcd, cad, abd; 1)$ , where  $a, b, c$  are co-prime in pairs, and that it is then the direct product of cyclic groups of orders  $abcd$  and  $d$ .

Consider the abelian group  $(l, m, n; 1)$  in the form

$$(2.21) \quad R^l = S^m = (RS)^n = 1, \quad RS = SR.$$

Let  $N$  denote the least common multiple of  $l$  and  $m$ . Then

$$(RS)^N = R^N S^N = 1;$$

so  $N$  must be a multiple of  $n$  (or the group would collapse). Thus, on account of the symmetry between  $l, m, n$ , each of these three members must divide the least common multiple of the other two.

Let  $mn/l = \lambda$ ,  $nl/m = \mu$ ,  $lm/n = \nu$ . Then  $l^2 = \mu\nu$ ,  $m^2 = \nu\lambda$ ,  $n^2 = \lambda\mu$ ; so we may write  $\lambda = a^2d$ ,  $\mu = b^2d$ ,  $\nu = c^2d$ , where  $a, b, c, d$  are positive integers. Thus

$$(2.22) \quad l = bcd, \quad m = cad, \quad n = abd.$$

By absorbing into  $d$  any common factor of  $a, b, c$ , we may suppose\* that  $(a, b, c) = 1$ .

Further, since  $N = abcd/(a, b)$  must be a multiple of  $n = abd$ ,  $c$  must be a multiple of  $(a, b)$ . Hence  $(a, b, c) = (a, b)$  and

$$(2.23) \quad (b, c) = (c, a) = (a, b) = 1.$$

We are thus led to consider the group  $R^{bcd} = S^{cad} = (RS)^{abd} = 1$ ,  $RS = SR$ , where  $a, b, c$  are co-prime in pairs. Accordingly, we may choose (positive or negative) integers  $\alpha, \beta, \gamma, \alpha', \beta', \gamma'$ , such that

$$\alpha b - \beta a = 1, \quad \gamma a - \alpha' c = 1, \quad \beta' c - \gamma' b = 1.$$

Then

$$\begin{aligned} (R^{\beta a} S^{\alpha b})^{\alpha b c d} &= S^{\alpha^2 b^2 c d} = S^{(\beta a + 1)^2 c d} = S^{c d}, \\ (R^{\beta a} S^{\alpha b})^{\beta c a d} &= R^{\beta^2 a^2 c d} = R^{(\alpha b - 1)^2 c d} = R^{c d}, \\ (R^{\beta a} S^{\alpha b})^{\gamma a b d} &= R^{(\beta a - \alpha b) \gamma a b d} = R^{-(\alpha' c + 1) b d} = R^{-b d}, \\ (R^{\beta a} S^{\alpha b})^{\gamma' a b d} &= S^{(\alpha b - \beta a) \gamma' a b d} = S^{(\beta' c - 1) a d} = S^{-a d}, \end{aligned}$$

\* Following the usual notation, we let  $(a, b, \dots)$  denote the greatest common factor of  $a, b, \dots$ .

and therefore

$$\begin{aligned} R^d &= R^{(\beta'c-\gamma'b)d} = (R^{\beta a} S^{ab})^{(\beta'\beta c+\gamma'\gamma b)ad}, \\ S^d &= S^{(\gamma a-\alpha'e)d} = (R^{\beta a} S^{ab})^{-(\gamma'\gamma a+\alpha'\alpha e)bd}, \\ (RS)^d &= R^d S^d = (R^{\beta a} S^{ab})^{(\beta'\beta a-\alpha'\alpha e)cd}. \end{aligned}$$

Since  $S = R^{\beta a} S^{ab} (RS)^{-\beta a}$ , it follows that every operator of the group is expressible in the form

$$(R^{\beta a} S^{ab})^z (RS)^y, \quad 0 \leq y < d,$$

and that the group is the direct product of cyclic groups, of orders  $abcd$  and  $d$ , generated by

$$R^{\beta a} S^{ab}, \quad (R^{\beta a} S^{ab})^{(\alpha'ab-\beta'\beta a)e} RS.$$

Thus, under the condition (2.23),

$$(2.24) \quad (bcd, cad, abd; 1) \sim [abcd]' \times [d]',$$

the product of two cyclic groups.

**2.3. The polyhedral groups and  $(2, 3, 7; q)$ .** When  $l=m=2$ , the last relation of (2.11) is  $(RS)^{2q}=1$ ; so  $(2, 2, n; q)$  collapses unless either  $n=2q$ , or  $n=q$  and  $n$  is odd. In the former case we have the dihedral group

$$(2.31) \quad (2, 2, 2q; q) \sim [2q],$$

and in the latter,

$$(2.311) \quad (2, 2, q; q) \sim [q], \quad q \text{ odd.}$$

The definition (2.13) enables us easily to evaluate  $q$  for the polyhedral groups  $[3, l]'$ , ( $l < 6$ ), regarded as  $(2, 3, l; q)$ . For the *tetrahedral group* ( $l=3$ ),  $R^3 T = T$ ; so  $q=2$ . For the *octahedral group* ( $l=4$ ),  $R^3 T = (TR)^{-1}$ ; so  $q=3$ . For the *icosahedral group* ( $l=5$ ),  $R^3 T = (TR^2)^{-1}$ ; so

$$(2.32) \quad (2, 3, 5; q) \sim (3, 5 | 2, q),$$

whence, by (1.16),  $q=5$ . Thus

$$(2.33) \quad (2, 3, 3; 2) \sim [3, 3]' \sim G_{41/2},$$

$$(2.34) \quad (2, 3, 4; 3) \sim [3, 4]' \sim G_{41},$$

$$(2.35) \quad (2, 3, 5; 5) \sim [3, 5]' \sim G_{51/2}.$$

In each case, any other value of  $q$  would cause collapse.

In the form  $(2, 3, 5; q) \sim (q, 5 | 2, 3)$  (see (1.17)), (2.32) remains true when

the fives are replaced by sevens, although the obvious generalization fails. For, writing  $S^2$  for  $R$  in (2.13) with  $l=7$ , we obtain

$$S^7 = T^2 = (S^2T)^3 = (ST)^2 = 1.$$

Thus

$$(2.36) \quad (2, 3, 7; q) \sim (q, 7 \mid 2, 3).$$

This provides an elegant proof for Brahana's result\* that  $(2, 3, 7; q)$  collapses when  $q=2, 3$ , or  $5$ . (We use (1.14), (1.16), and (1.18), respectively.) His definition†

$$Q^2 = R^7 = (QR^3)^2 = (QR^2)^3 = 1$$

is equivalent to (2.13) ( $l=7$ ) with  $T=QR^3$ .

By (2.36) and (1.66),

$$(2.37) \quad (2, 3, 7; 4) \sim \mathfrak{P}_1(7).$$

In the form

$$A_1^7 = A_2^3 = (A_1A_2)^2 = (A_2A_1)^5 = 1$$

(see (2.12)), this definition was given by Dyck.‡ Thus (2.36) (with  $q=4$ ) relates Dyck's definition to Burnside's.

By (2.36) and (1.72), we have

$$(2.38) \quad (2, 3, 7; 6) \sim (2, 3, 7; 7) \sim \mathfrak{P}_1(13).$$

These definitions are due to Brahana and Sinkov, respectively.§ By (1.75) and (1.751),  $(2, 3, 7; 6)$  is generated by

$$(2.39) \quad \begin{cases} R = (0 \ 10 \ 2 \ 9 \ 3 \ 8 \ 5)(1 \ 7 \ 11 \ 4 \ 6 \ \infty \ 12), \\ T = (1 \ 12)(2 \ 11)(3 \ 10)(4 \ 9)(5 \ 8)(6 \ 7); \end{cases}$$

and by (1.752),  $(2, 3, 7; 7)$  is generated by the same  $T$  and

$$(2.391) \quad R = (0 \ 9 \ 11 \ 5 \ 12 \ 1 \ 10)(2 \ 7 \ 6 \ 4 \ 3 \ 8 \ \infty).$$

Sinkov has shown|| that  $(2, 3, 7; 8)$  has a factor group of order 10752. When written as a factor group of  $(8, 7 \mid 2, 3)$ , this takes the form

$$R^8 = S^7 = (RS)^2 = (R^{-1}S)^3 = (R^2S^4)^6 = 1.$$

**2.4. A lemma on automorphisms.** We shall often have occasion to use the following general principle, which the reader will have no difficulty in proving.

\* Brahana [2], pp. 350, 352.

† Ibid., p. 349; quoted by Sinkov [3], p. 68.

‡ Dyck [1], p. 41; Burnside [1], p. 422.

§ Brahana [2], p. 354; Sinkov [1], p. 239.

|| Sinkov [5].

If a group  $G$  is augmented by the adjunction of an operator  $T$ , of period  $n$ , which transforms  $G$  according to an *inner* automorphism of the same period, and if the order of the central of  $G$  is prime to  $n$ , then the augmented group is the direct product  $G \times \{T\}$ .

The necessity\* for mentioning the central is illustrated by the following example, due to R. Brauer. Let  $G$  be the trigonal dicyclic group† (of order 12),  $B^3 = C^4 = 1$ ,  $C^{-1}BC = B^{-1}$ , with generators  $B = (1\ 2\ 3)$ ,  $C = (2\ 3)(4\ 5\ 6\ 7)$ . Then the operator  $T = (2\ 3)$  transforms  $G$  in the same manner as  $C$ . But  $\{B, C, T\}$ , being generated by  $(1\ 2\ 3)$ ,  $(2\ 3)$ , and  $(4\ 5\ 6\ 7)$ , is the direct product of the symmetric group of order six and the cyclic group of order four (*not* the direct product of the dicyclic group and the group of order two).

**2.5. The derivation of  $(2, m, m; q)$  and  $(2, m, 2n; k)$  from  $(4, m|2, 2q)$  and  $(m, m|n, k)$ .** The group  $R^m = S^m = (RS)^n = (R^2S^2)^q = (R^{-1}S)^k = 1$  has an automorphism which interchanges the generators. We adjoin an involutory operator  $T$ , which transforms the group according to this automorphism. Substituting  $TST$  for  $R$ , we obtain the augmented group in the form‡  $S^m = T^2 = (ST)^{2n} = (S^2T)^{2q} = (S^{-1}TST)^k = 1$ . This result provides two useful lemmas, the first by ignoring  $k$  and putting  $n=2$ , the second by ignoring  $q$ :

(2.51)  $(2, m, m; q)$  is a subgroup of index two in  $(4, m|2, 2q)$ .

(2.52)  $(2, m, 2n; k)$  contains  $(m, m|n, k)$  as a subgroup of index two.

Since  $(m, m|n, k) \sim (m, m|k, n)$ , it follows that the two groups  $(2, m, 2n; k)$ ,  $(2, m, 2k; n)$  have the same order. Since  $(m, m|k, n)$  is derived from  $(m, m|n, k)$  by writing  $R^{-1}$  for  $R$ , we may say that  $(2, m, 2k; n)$  is derived from  $(m, m|n, k)$  by adjoining an involutory operator  $T'$ , which transforms  $S$  into  $R^{-1}$  (and  $R$  into  $S^{-1}$ ).

In particular,§ since  $(4, 4|2, k)$  is of order  $4k^2$ , both  $(2, 4, 4; k)$  and  $(2, 4, 2k; 2)$  are of order  $8k^2$ . The order of  $(2, 4, 2k; 2)$  can alternatively be deduced from that of its subgroup  $(2k, 2k|2, 2)$  which was described in §1.3. Again,||  $(2, 3, 6; k)$  and  $(2, 3, 2k; 3)$  are of order  $6k^2$ , since both contain  $(3, 3|3, k)$  as a subgroup of index two.

By (2.51),  $(2, 4, 4; k)$  is a subgroup of index two in  $(4, 4|2, 2k)$ , and  $(2, 5, 5; 2)$  in  $(4, 5|2, 4)$ . By (1.42), it follows that¶  $(2, 5, 5; 2)$ , of order eighty, is generated by

\* Pointed out by a referee.

† In terms of  $BC^2$  and  $C$ , this group has the simpler definition  $A^3 = C^2 = (AC)^2$ .

‡ That is,  $(2n, m|2, 2q; k)$ .

§ Edington [1], p. 197; Sinkov [2], p. 79.

|| Edington [1], p. 207; Sinkov [2], p. 82.

¶ Coxeter [6], p. 284.

$$(2.53) \quad S = (0 \ 1 \ 2 \ 3 \ 4)(5 \ 6 \ 7 \ 8 \ 9), \quad TST = (0 \ 9 \ 8 \ 7 \ 6)(5 \ 4 \ 3 \ 2 \ 1).$$

A number of special applications of (2.52) may conveniently be arranged as columns of a table:

$(m, m   n, k)$	$(5, 5   2, 3) \sim G_{51/2}$	$(3, 3   4, 4) \sim \mathfrak{P}_1(7)$
$R$ $S$	$(2 \ 1 \ 3 \ 4 \ 5)$ $(1 \ 2 \ 3 \ 4 \ 5)$	$(0 \ 1 \ \infty)(2 \ 6 \ 4)$ $(0 \ 2 \ 3)(1 \ 6 \ 5)$
Automorphism $(R, S)$	$(1 \ 2)$ (outer)	$(1 \ 2)(3 \ \infty)(4 \ 5)$ (outer)
Automorphism $(R^{-1}, S)$	$(3 \ 5)$ (outer)	
$(2, m, 2n; k)$ $(2, m, 2k; n)$	$(2, 4, 5; 3) \sim G_{51}^*$ $(2, 5, 6; 2) \sim G_{51}$	$(2, 3, 8; 4) \sim \mathfrak{P}_1(7)$

$(m, m   n, k)$	$(5, 5   2, 4) \sim G_{51/2}$	$(7, 7   2, 3) \sim \mathfrak{P}_1(13)$
$R$ $S$	$(1 \ 6 \ 5 \ 4 \ 3)\dagger$ $(1 \ 2 \ 3 \ 4 \ 5)$	$(0 \ 5 \ 7 \ 8 \ 10 \ 2 \ 1)(3 \ 11 \ 4 \ 12 \ 9 \ \infty \ 6)$ $(5 \ 0 \ 12 \ 9 \ 1 \ 11 \ 10)(4 \ 2 \ 3 \ 7 \ 8 \ 6 \ \infty)$
Automorphism $(R, S)$	$(2 \ 6)(3 \ 5)$ (inner)	$(0 \ 5)(1 \ 10)(2 \ 11)(3 \ 4)(6 \ \infty)(7 \ 12)(8 \ 9)$ (outer)
Automorphism $(R^{-1}, S)$	(degree greater than 6) (outer)	$(0 \ \infty)(1 \ 4)(3 \ 10)(5 \ 6)(7 \ 8)(9 \ 12)$ (inner) $\ddagger$
$(2, m, 2n; k)$ $(2, m, 2k; n)$	$(2, 4, 5; 4) \sim G_{51/2} \times G_2$ $(2, 5, 8; 2)$ (see below)	$(2, 4, 7; 3) \sim \mathfrak{P}_1(13)$ $(2, 6, 7; 2) \sim \mathfrak{P}_1(13) \times G_2$

\* Burnside [1], p. 422.

$\dagger$  Todd and Coxeter [1], p. 31.

$\ddagger$  Since  $S^2R^{-2} = (0 \ 4 \ \infty \ 8 \ 12 \ 10 \ 2)(1 \ 7 \ 9 \ 6 \ 3 \ 5 \ 11)$ , the relations  $S^7 = T'^2 = (ST')^6 = (S^{-1}T'ST')^2 = 1$  imply  $(S^2T')^{14} = 1$ . Hence  $(6, 7 | 2, 7; 2) \sim \mathfrak{P}_1(13)$ .

In each of these cases,  $(m, m | n, k)$  is a simple group; so the order of its central is 1. The interchange of  $R$  and  $S$  (or of  $R^{-1}$  and  $S$ ) is usually recognizable as an inner or outer automorphism according as it is an even or odd permutation.

In describing  $(2, 3, 8; 4)$  and  $(2, 4, 7; 3)$ , we appeal to the theorem (of Schreier and van der Waerden) that the group of isomorphisms of  $\mathfrak{P}_1(p)$  ( $p$  prime) is  $\tilde{\mathfrak{P}}_1(p)$ . It is the same theorem which enables us to assert that the even permutation  $(0 \infty)(1 4)(3 10)(5 6)(7 8)(9 12)$  transforms  $\mathfrak{P}_1(13)$  according to an inner automorphism.

In the case of  $(5, 5 | 2, 4)$ , since  $R^{-1}$  and  $S$  are not interchanged by any permutation of degree six,  $T'$  transforms  $G_{61/2}$  according to an outer automorphism of  $G_6$ . In other words,  $(2, 5, 8; 2)$  (of order 720) is the nonsymmetric subgroup of index two in the group of isomorphisms of  $G_6$ . It is easily verified that the following permutations of degree ten satisfy the defining relations for  $(5, 5 | 2, 4)$  and so generate  $G_{61/2}$ :

$$(2.54) \quad \begin{aligned} R &= (4 \ 7 \ 8 \ 3 \ 0)(9 \ 6 \ 1 \ 2 \ 5), \\ S &= (0 \ 1 \ 2 \ 3 \ 4)(5 \ 6 \ 7 \ 8 \ 9). \end{aligned}$$

In this representation,  $R^{-1}$  and  $S$  are interchanged by

$$T' = (0 \ 5)(1 \ 2)(3 \ 6)(4 \ 9)(7 \ 8).$$

Hence  $(2, 5, 8; 2)$  is generated by the permutations  $S$  and  $T'$ .

Finally, let us apply (2.52) to Miller's group  $(3, 3 | 4, 5)$ , of order 1080, so as to derive  $(2, 3, 8; 5)$  and  $(2, 3, 10; 4)$ , of order 2160. Since, in (1.74), the permutation  $R$  involves six cycles, while  $S$  involves only five, the transformations effected by  $T$  and  $T'$  are outer automorphisms of  $(3, 3 | 4, 5)$ . But the combined operator  $TT'$ , which transforms  $R$  and  $S$  into their own inverses, behaves like the permutation

$$(2.55) \quad P = (2 \ 3)(5 \ 6)(7 \ 8)(10 \ 11)(13 \ 14)(16 \ 17).$$

The generators

$$R_1 = (1 \ 2 \ 3)(4 \ 5 \ 6), \quad S_1 = (2 \ 3 \ 4),$$

of the central quotient group of  $(3, 3 | 4, 5)$ , are transformed into their own inverses by

$$(2 \ 3)(5 \ 6) = (R_1 S_1)^2 (S_1 R_1)^2 (R_1 S_1)^2 (S_1 R_1)^2 (R_1 S_1)^2.*$$

It is easily verified that

$$(RS)^2 (SR)^2 (RS)^2 (SR)^2 (RS)^2 = P.$$

Hence the transformation effected by  $TT'$  is an inner automorphism of  $(3, 3 | 4, 5)$ , and

$$(2.56) \quad (2, 3, 8; 5) \sim (2, 3, 10; 4).$$

\* For this expression I am indebted to Dr. Sinkov.

By (2.13),  $(2, 3, l; q)$  is the group of the skew polyhedron\*  $\{3, l | , q\}$ . Hence the two polyhedra  $\{3, 8 | , 5\}$  and  $\{3, 10 | , 4\}$  have the same group, as was implied in Table II† at the end of *Regular skew polyhedra in three and four dimensions, and their topological analogues*.

We have already remarked (1.741) that  $(3, 3 | 4, 5)$  has a central, of order three, generated by‡  $Z = (R^{-1}S^{-1}RS)^5$ . Since  $T$  (and likewise  $T'$ ) transforms this operator into its inverse, the group of order three generated by  $Z$ , regarded as a subgroup of  $(2, 3, 8; 5)$  or  $(2, 3, 10; 4)$ , is no longer central although of course it is still invariant. Its quotient group, of order 720, is derived from  $G_{61/2}$  by adjoining an operator which transforms the latter according to an outer automorphism of  $G_{61}$ , that is,

$$(2.57) \quad (2, 3, 8; 5)/G_3 \sim (2, 5, 8; 2).$$

In fact, the generators of  $G_{61/2}$ , as permutations of degree ten, take the form of

$$R_1 = (0\ 2\ 7)(1\ 8\ 4)(5\ 6\ 9), \quad S_1 = (9\ 7\ 2)(8\ 1\ 5)(4\ 3\ 0).$$

These are interchanged by

$$T_1 = (0\ 9)(1\ 8)(2\ 7)(3\ 6)(4\ 5).$$

It is easily verified that  $S_1$  and  $T_1$  satisfy the relations

$$S_1^3 = T_1^2 = (S_1 T_1)^8 = (S_1^{-1} T_1 S_1 T_1)^5 = 1.$$

**2.6. The criterion for finiteness.** By applying (2.51) and (2.52) to the general groups considered in Theorem A, we deduce the following theorem:

**THEOREM C.** *If  $q > 1$  and  $1/m + 1/n \leq 1/2$ , and if  $m$  and  $n$  are either even or equal (or both), the group  $(2, m, n; q)$  is finite when*

$$(2.61) \quad \cos 2\pi/m + \cos 2\pi/n + \cos \pi/q < 1,$$

*and infinite otherwise.*

For example,  $(2, 4, 6; 3)$ ,  $(2, 6, 6; 2)$ , and  $(2, 5, 5; 3)$  are infinite. The condition  $1/m + 1/n \leq 1/2$  is inserted in order to exclude the polyhedral groups of §2.3, for which  $q$  is a function§ of  $m, n$ .

The groups  $(2, 4, 2k; 2)$ ,  $(2, 4, 4; q)$ , and  $(2, 5, 5; 2)$  are the only finite groups which satisfy all the conditions of Theorem C. If we allow  $m, n$  to be

\* Coxeter [7], p. 59.

† Ibid., p. 61.

‡ Cf. Miller [1], p. 365, where  $Z$  is expressed in the form  $S\{(S^{-1}R)^2(SR^{-1})^2\}^2S^{-1}$ .

§ See (4.71), below.

unequal and (one or both) odd, we find that the criterion (2.61) admits the groups

$$\begin{aligned} (2, 3, 2k; 3), \quad (2, 3, 6; q), \quad (2, 3, 7; 4), \\ (2, 3, 7; 5) \text{ (collapsing)}, \quad (2, 3, 7; 6), \\ (2, 3, 8; 4), \quad (2, 4, 5; 3), \quad (2, 5, 6; 2), \end{aligned}$$

which we have already discussed, and admits also the following further possibilities:

$$\begin{aligned} (2, 3, n; 2), \quad (2, 3, 9; 4), \quad (2, 5, 7; 2), \\ (2, 3, n; 3) \text{ (} n \text{ odd)}, \quad (2, 4, n; 2) \text{ (} n \text{ odd)}. \end{aligned}$$

Sinkov has proved\* that  $(2, 3, n; 2)$  for  $n \neq 3, 6$ ,  $(2, 4, n; 2)$  for  $n$  odd, and  $(2, 3, n; 3)$  for  $n$  odd, all collapse. Since  $(2, 4, n; 2)$  contains  $(n, n | 2, 2)$  as a subgroup of index two, the collapse of  $(2, 4, n; 2)$  ( $n$  odd) can alternatively be deduced from (1.19). For the collapse of  $(2, 3, 9; 4)$  and  $(2, 5, 7; 2)$ , see the Appendix; in both cases the enumeration of cosets by the Todd-Coxeter method soon breaks down.

These results show that  $(2, m, n; q)$  is finite (or collapses) whenever (2.61) is satisfied. Hence we have the following theorem:

**THEOREM D.** *For all infinite groups  $(2, m, n; q)$ ,*

$$(2.62) \quad \cos 2\pi/m + \cos 2\pi/n + \cos \pi/q \geq 1.$$

We saw, in §2.2 and §2.4, that the groups

$$(2, 3, 7; 7), (2, 3, 8; 5) \sim (2, 3, 10; 4), (2, 4, 5; 4), (2, 4, 7; 3), (2, 5, 8; 2), (2, 6, 7; 2)$$

are finite, although they satisfy (2.62) (see §1.7). There are at least two further cases of this kind. By enumerating the 380 cosets of the subgroup of order nine generated by  $S$ , I found the order of  $(2, 5, 9; 2)$  to be 3420. Since the defining relations

$$S^9 = T^2 = (ST)^5 = (S^{-1}TST)^2 = 1$$

are satisfied by the linear fractional substitutions

$$S = (-3x), \quad T = \left( \frac{x-7}{x-1} \right) \pmod{19},$$

we can therefore assert that

$$(2.63) \quad (2, 5, 9; 2) \sim \mathbb{P}_1(19).$$

\* Sinkov [2], pp. 75, 79, 82.

As permutations, we have

$$(2.64) \quad \begin{cases} S = (1 \ 16 \ 9 \ 11 \ 5 \ 4 \ 7 \ 17 \ 6)(2 \ 13 \ 18 \ 3 \ 10 \ 8 \ 14 \ 15 \ 12), \\ T = (0 \ 7)(1 \ \infty)(2 \ 14)(3 \ 17)(4 \ 18)(5 \ 9)(6 \ 15)(8 \ 11)(10 \ 13)(12 \ 16). \end{cases}$$

Again, by enumerating\* the 552 cosets of the subgroup of order eleven generated by  $S$ , I found the order of  $(2, 3, 11; 4)$  to be 6072. Since the defining relations

$$S^{11} = T^2 = (ST)^3 = (S^{-1}TST)^4 = 1$$

are satisfied by the linear fractional substitutions†

$$S = (9x), \quad T = \left( \frac{x-3}{x-1} \right) \pmod{23},$$

we can therefore assert that

$$(2.65) \quad (2, 3, 11; 4) \sim \mathfrak{P}_1(23).$$

As permutations, we have

$$(2.66) \quad \begin{cases} S = (1 \ 9 \ 12 \ 16 \ 6 \ 8 \ 3 \ 4 \ 13 \ 2 \ 18)(5 \ 22 \ 14 \ 11 \ 7 \ 17 \ 15 \ 20 \ 19 \ 10 \ 21), \\ T = (0 \ 3)(1 \ \infty)(2 \ 22)(4 \ 8)(5 \ 12)(6 \ 19)(7 \ 16)(9 \ 18)(10 \ 11)(13 \ 20)(14 \ 15)(17 \ 21). \end{cases}$$

This result adds one to the list of *finite polyhedra*  $\{l, m, q\}$ ,‡ namely  $\{3, 11, 4\}$ , with 2024 (triangular) faces, 3036 edges, 552 vertices, genus 231, and group  $LF(2, 23)$ , of order 6072.

On the other hand, Theorem C does not cover *all* the infinite groups  $(2, m, n; q)$ . For, Brahana§ has shown that  $(2, 3, 8; 6)$  is infinite, whence, by (2.52),

$$(2.67) \quad (3, 3 \mid 4, 6), \quad (2, 3, 12; 4) \text{ are infinite.}$$

Fig. 4 (on p. 121) shows a graphical enumeration of groups  $(2, n, p; 2)$ . Known finite groups are represented by o's, known infinite groups by dots, and known cases of collapse by crosses.

**2.7. The derivation of  $(n, n, n; q)$ ,  $(3, 3, 3; k)$ ,  $(3, 3, 4; 3)$  from  $(3, 3 \mid n, 3q)$ ,  $(3, 3 \mid 3, k)$ ,  $(3, 3 \mid 4, 4)$ .** The group

$$R^n = S^n = T^n = RST = (TSR)^q = (R^{-1}S)^k = (S^{-1}T)^k = (T^{-1}R)^k = 1$$

\* After some practice, such enumeration proceeds at the rate of about sixty cosets per hour.

† Cf. the substitutions  $S = (4x)$ ,  $T = [(x-5)/(x-1)] \pmod{23}$ , which satisfy  $S^{11} = T^2 = (ST)^4 = (S^2T)^3 = 1$ . However, the order of  $(4, 11 \mid 2, 3)$  is certainly greater than 6072; so  $\mathfrak{P}_1(23)$  is a proper factor group.

‡ Coxeter [7], p. 61, Table II.

§ Brahana [4], p. 901.

has an automorphism which cyclically permutes the three generators. We adjoin an operator  $Q^{-1}$ , of period three, which transforms the group according to this automorphism. Since  $T = QSQ^{-1}$  and  $R = Q^{-1}SQ$ , the augmented group is defined by\*

$$Q^3 = S^n = (QS)^3 = (Q^{-1}S)^{3q} = (S^{-1}Q^{-1}SQ)^k = 1.$$

By ignoring first  $k$  and then  $q$ , we deduce two lemmas analogous to (2.51) and (2.52):

(2.71)  $(n, n, n; q)$  is an invariant subgroup of index three in  $(3, 3 | n, 3q)$ .

(2.72)  $(3, 3, n; k)$  contains

$$R^n = S^n = T^n = RST = (R^{-1}S)^k = (S^{-1}T)^k = (T^{-1}R)^k = 1$$

as an invariant subgroup of index three.

This last group reduces to something familiar in two special cases. When  $n=3$ , since

$$S^{-1}T = S^2T = SR^{-1}, \quad RT^{-1} = R^{-2}T^{-1} = R^{-1}S,$$

we are left with

$$R^3 = S^3 = (RS)^3 = (R^{-1}S)^k = 1.$$

Since  $(3, 3 | 3, k)$  is of order  $3k^2$ , it follows that  $(3, 3, 3; k)$  is of order  $9k^2$ .

Secondly, when  $n=4$  and  $k=3$ , since

$$\begin{cases} SR^{-1} = S^2T, \\ R^{-1}T = ST^2 = ST^{-1} \cdot T^{-1} = (TS^{-1})^2 T^{-1} = TS \cdot S^2T \cdot S^{-1}T^{-1}, \\ (S^2T)^2 R \cdot S^2T = S^2TS^{-1}T = S^2 \cdot ST^{-1}S = S^{-1}T^{-1}S, \end{cases}$$

we are left with

$$S^4 = T^4 = (S^{-1}T)^3 = (S^2T)^3 = 1$$

(compare (1.12) with  $S^{-1}$  for  $S$ ). Since  $(3, 4 | 3, 4) \sim \mathfrak{P}_1(7)$ , it follows that  $(3, 3, 4; 3)$  is of order 504. We easily verify that  $\mathfrak{P}_1(7)$ , in the form

$$R^4 = S^4 = T^4 = RST = (R^{-1}S)^3 = (S^{-1}T)^3 = (T^{-1}R)^3 = 1,$$

is generated by the permutations

$$R = (1\ 3)(2\ 4\ 6\ 5), \quad S = (2\ 7)(4\ 1\ 3\ 5), \quad T = (4\ 6)(1\ 2\ 7\ 5),$$

which are cyclically permuted by  $(1\ 2\ 4)(3\ 7\ 6)$ . Since this transforms the

\* That is,  $(3, 3 | n, 3q; k)$ .

simple group  $\mathfrak{P}_1(7)$  according to an inner automorphism,  $(3, 3, 4; 3)$  is a direct product,\* namely

$$(2.73) \quad (3, 3, 4; 3) \sim \mathfrak{P}_1(7) \times G_3.$$

Putting  $q=1$  in (2.71), we see that the group  $(3, 3 | 3, n)$ , of order  $3n^2$ , can be derived from

$$R^n = S^n = T^n = RST = TSR = 1$$

(which, by (2.24), is the direct product of two cyclic groups of order  $n$ ) by adjoining an operator of period three which cyclically permutes  $R, S, T$ . (When  $n=2$ , this is the familiar derivation of the tetrahedral group from the four-group.) As permutations, we may write

$$\begin{aligned} R &= (1\ 4\ 7 \cdots 3n-2)(3n\ 3n-3 \cdots 6\ 3), \\ S &= (2\ 5\ 8 \cdots 3n-1)(3n-2\ 3n-5 \cdots 4\ 1), \\ T &= (3\ 6\ 9 \cdots 3n)(3n-1\ 3n-4 \cdots 5\ 2); \end{aligned}$$

whence

$$Q^{-1} = (1\ 2\ 3)(4\ 5\ 6) \cdots (\cdots 3n).$$

The permutations (1.63), which generate  $(3, 3 | 3, k)$ , are just  $Q^{-1}$  and  $Q^{-1}R$  (with  $k$  in place of  $n$ ).

By (2.71) again,  $(3, 3, 3; q)$  is a subgroup of index three in  $(3, 3 | 3, 3q)$ , and  $(4, 4, 4; 2)$  in  $(3, 3 | 4, 6)$ . Hence, by (2.67),

$$(2.74) \quad (4, 4, 4; 2) \text{ is infinite.}$$

**2.8. Groups of genus one.**† The infinite group  $[4, 4]'$ , defined by

$$R^4 = S^4 = (RS)^2 = 1,$$

has, as is well known,‡ an abelian subgroup of index four, generated by  $RS^{-1}$  and  $R^{-1}S$ . When represented in the usual way in the euclidean plane, these operators appear as translations. Hence every operator of  $[4, 4]'$  is expressible in the form

$$S^n(RS^{-1})^p(R^{-1}S)^q.$$

The most general factor group is given by

$$(2.81) \quad (RS^{-1})^p(R^{-1}S)^q = 1,$$

which takes the form  $(RS^{-1}R^{-1}S)^q$  or  $(R^2S^2)^q = 1$  when  $p=q$ . The relation

\* Since  $(Q^{-1}S)^3 = RTS = (1\ 5\ 2\ 6\ 3\ 7\ 4)$ , the relations  $Q^3 = S^3 = (QS)^3 = (S^{-1}Q^{-1}SQ)^3 = 1$  imply  $(Q^{-1}S)^3 = 1$ . Hence  $(3, 3 | 4, 7; 3) \sim (3, 4 | 3, 7; 3) \sim \mathfrak{P}_1(7)$ .

† Cf. Sinkov [4].

‡ Burnside [1], p. 416.

$(R^{-1}S)^q = 1$  implies  $(R^2S^2)^q = 1$ , while  $(R^2S^2)^q = 1$  implies  $(R^{-1}S)^{2q} = 1$ . In fact,  $^*(4, 4 | 2, q)$  is the central quotient group of  $(2, 4, 4; q)$ , and  $(2, 4, 4; q)$  is the central quotient group of  $(4, 4 | 2, 2q)$ . By (2.52) and (2.51), these relationships are retained when we write "a subgroup" in place of "the central quotient group."

By (2.51) and (1.32),  $(2, 4, 4; q)$  is generated by  $R^{-1}SR$  and  $S$  (or by  $R$  and  $S^{-1}RS$ ),<sup>†</sup> where

$$(2.82) \quad \begin{cases} R = (1 \ 2 \ 3 \ 4)(5 \ 6 \ 7 \ 8) \cdots (\cdots 4q), \\ S = (1 \ 2)(3 \ 4 \ 5 \ 6) \cdots (\cdots 4q - 2)(4q - 1 \ 4q). \end{cases}$$

When  $k$  is odd, (1.32) shows that  $(4, 4 | 2, k)$  is generated by the permutations

$$(2.83) \quad \begin{cases} R = (1 \ 2 \ 3 \ 4) \cdots (\cdots 2k - 2)(2k - 1 \ 2k), \\ S = (1 \ 2)(3 \ 4 \ 5 \ 6) \cdots (\cdots 2k), \end{cases}$$

which are interchanged by

$$T_1 = (1 \ 2k - 1)(2 \ 2k)(3 \ 2k - 3)(4 \ 2k - 2) \cdots (k - 2 \ k + 2)(k - 1 \ k + 3).$$

Since

$$T_1 S^2 = (1 \ 2k - 3 \ 5 \ 2k - 7 \cdots 7 \ 2k - 5 \ 3 \ 2k - 1)(2 \ 2k - 2 \ 6 \ 2k - 6 \cdots 8 \ 2k - 4 \ 4 \ 2k),$$

which is of period  $k$ ,  $S$  and  $T_1$  satisfy the defining relations of  $(4, 4 | 2, k)$ ; that is,  $T_1$  transforms  $(4, 4 | 2, k)$  according to an inner automorphism. In fact,

$$T_1 = (R^2 S^2)^{(k-1)/2} R^2.$$

Now,  $k$  being odd, the central of  $(4, 4 | 2, k)$  is of order 1. Hence  $(2, 4, 4; k)$  is a direct product, namely

$$(2.84) \quad (2, 4, 4; k) \sim (4, 4 | 2, k) \times G_2, \quad k \text{ odd}.$$

We proceed to find a representation for  $(2, 4, 2k; 2)$  ( $k$  unrestricted). Formula (1.31) shows that  $(2k, 2k | 2, 2)$  is generated by the permutations

$$(2.85) \quad \begin{cases} R = (a_1 \ a_2 \cdots a_{2k})(b_1 \ b_{2k})(b_2 \ b_{2k-1}) \cdots (b_k \ b_{k+1}), \\ S = (b_1 \ b_2 \cdots b_{2k})(a_1 \ a_{2k})(a_2 \ a_{2k-1}) \cdots (a_k \ a_{k+1}), \end{cases}$$

which are interchanged by

$$(2.851) \quad T = (a_1 \ b_1)(a_2 \ b_2) \cdots (a_{2k} \ b_{2k}).$$

Since both  $R$  and  $S$  permute the  $a$ 's among themselves,  $T$  transforms

\* Sinkov [4], p. 169; Edington [1], p. 201.

† Edington (loc. cit., p. 198) gives a representation for  $(4, 4 | 2, \alpha)$ , of degree  $4\alpha$ . (Ours is of degree  $2\alpha$ .) He describes (p. 203) the group  $(2, 4, 4; \beta)$ , of order  $8\beta^2$ , but gives no representation.

$(2k, 2k | 2, 2)$  according to an outer automorphism. Therefore  $S$  and  $T$  generate  $(2, 4, 2k; 2)$ . (Other permutations of the same degree were given by Sinkov.\*) When  $k$  is even, there is a central of order two which is generated by

$$(TS^k)^2 = R^k S^k = (a_1 a_{k+1}) \cdots (a_k a_{2k}) (b_1 b_{k+1}) \cdots (b_k b_{2k}).$$

When  $k$  is odd, we can obtain a representation of degree  $2k$  (instead of  $4k$ ), by deriving  $(2, 4, 2k; 2)$  from  $(4, 4 | 2, k)$ . Using the generating permutations (2.83) for  $(4, 4 | 2, k)$  ( $k$  odd), we seek a permutation which will transform  $R, S$  into  $S^{-1}, R^{-1}$ . Such is clearly

$$(2.86) \quad T' = (1 \ 2k)(2 \ 2k-1) \cdots (k \ k+1).$$

This transforms  $(4, 4 | 2, k)$  according to an outer automorphism, since  $ST'$  is of period  $2k$  although  $(4, 4 | 2, k)$  contains no operator whose period is greater than  $k$ . Hence  $(2, 4, 2k; 2)$  ( $k$  odd) is generated by  $T'$  and

$$(2.861) \quad S = (1 \ 2)(3 \ 4 \ 5 \ 6) \cdots (\cdots 2k).$$

The infinite group

$$R^3 = S^3 = (RS)^3 = 1$$

is closely analogous to  $[4, 4]'$ . It has† an abelian subgroup of index three, generated by  $RS^{-1}$  and  $R^{-1}S$ . When represented in the usual way in the euclidean plane, these operators appear as translations. The most general factor group is again given by (2.81), which takes the form  $(RS^{-1}R^{-1}S)^q = 1$  when  $p=q$ . The relation  $(R^{-1}S)^q = 1$  implies  $(RS^{-1}R^{-1}S)^q = 1$ , while  $(RS^{-1}R^{-1}S)^q = 1$  implies  $(R^{-1}S)^{3q} = 1$ . In fact,‡  $(3, 3 | 3, q)$  is the central quotient group of  $(3, 3, 3; q)$ , and  $(3, 3, 3; q)$  is the central quotient group of  $(3, 3 | 3, 3q)$ . By (2.72) and (2.71), these relationships are retained when we write "a subgroup" in place of "the central quotient group."

By (2.71) ( $n=3$ ) and (1.63),  $(3, 3, 3; q)$  is generated by  $QSQ^{-1}$  and  $S$ , where

$$Q = (1 \ 2 \ 3)(4 \ 5 \ 6) \cdots (9q-2 \ 9q-1 \ 9q),$$

$$S = (9q \ 1 \ 2)(3 \ 4 \ 5) \cdots (9q-3 \ 9q-2 \ 9q-1).$$

In the form

$$R^3 = S^3 = T^3 = RST = (R^{-1}S)^k = (S^{-1}T)^k = (T^{-1}R)^k = 1,$$

$(3, 3 | 3, k)$  is generated by

\* Sinkov [2], p. 79.

† Burnside [1], p. 414.

‡ Sinkov [4], p. 168; Edington [1], p. 210.

$$(2.87) \quad \begin{cases} R = (1\ 2\ 3)(4\ 5\ 6) \cdots (3k-2\ 3k-1\ 3k), \\ S = (3k\ 1\ 2)(3\ 4\ 5) \cdots (3k-3\ 3k-2\ 3k-1), \\ T = (3k-1\ 3k\ 1)(2\ 3\ 4) \cdots (3k-4\ 3k-3\ 3k-2). \end{cases}$$

When  $k \equiv 0 \pmod{3}$ , there is no permutation of  $1, 2, \dots, 3k$  which will cyclically permute  $R, S, T$ . When  $k \equiv \pm 1 \pmod{3}$ ,  $R, S, T$  are permuted by  $Q_1^{-1}$ , where

$$Q_1^{\pm 1} = (1\ k+1\ 2k+1)(2\ k+2\ 2k+2) \cdots (k\ 2k\ 3k).$$

Clearly

$$Q_1^{-1}S = (1\ 2k+2\ k+3\ 4 \cdots 2k-1\ k)(k+1\ 2\ 2k+3\ k+4 \cdots 3k-1\ 2k)(k-1\ 3k-3\ 2k-5 \cdots k+5\ 3\ 2k+1)$$

or

$$(1\ k+2\ 2k+3\ 4 \cdots k-1\ 2k)(2k+1\ 2\ k+3\ 2k+4 \cdots 3k-1\ k)(2k-1\ 3k-3\ k-5 \cdots 2k+5\ 3\ k+1)$$

according as  $k \equiv 1$  or  $-1 \pmod{3}$ . Since this is of period  $k$ ,  $Q_1$  and  $S$  satisfy the defining relations of  $(3, 3 | 3, k)$ ; that is,  $Q_1$  transforms  $(3, 3 | 3, k)$  according to an inner automorphism. Since  $k$  is not divisible by 3, the central of  $(3, 3 | 3, k)$  is of order 1. Hence  $(3, 3, 3; k)$  is a direct product,\* namely

$$(2.88) \quad (3, 3, 3; k) \sim (3, 3 | 3, k) \times G_3, \quad k \not\equiv 0 \pmod{3}.$$

The generators  $R, S$  of  $(3, 3 | 3, k)$  ((1.63) or (2.87)) are interchanged by either of the permutations

$$T = (3\ 3k)(4\ 3k-2)(5\ 3k-1)(6\ 3k-3)(7\ 3k-5)(8\ 3k-4) \cdots,$$

$$T_1 = (1\ 3k-2)(2\ 3k-1)(3\ 3k-3)(4\ 3k-5)(5\ 3k-4)(6\ 3k-6) \cdots.$$

The permutation  $T$  transforms  $(3, 3 | 3, k)$  according to an outer automorphism,† since  $ST$  is of period six although  $(3, 3 | 3, k)$  contains, in general, no operator of that period. Hence, by (2.52),  $(2, 3, 6; k)$  is generated‡ by  $T$ , and

$$(2.89) \quad S = (3k\ 1\ 2)(3\ 4\ 5) \cdots (\cdots 3k-1).$$

$T$  can be replaced by  $T_1$  whenever  $k > 2$ ; but when  $k = 2$ ,  $ST_1$  is of period three instead of six, showing that  $(2, 3, 6; 2)$  is the direct product of the tetrahedral group and the group of order two, that is, the *pyritohedral* group.

\* See Miller [2], p. 668, for the case when  $k=2$ .

† This can also be seen by trying to express  $T$  in the form  $S^a(RS^{-1})^p(R^{-1}S)^q$ .

‡ Cf. Edington [1], p. 205, where a representation of degree  $6k$  is given.

On the other hand,  $R^{-1}$  and  $S$  are interchanged by

$$(2.891) \quad T' = (1 \ 3k - 1)(2 \ 3k - 2)(3 \ 3k - 3) \cdots.$$

This and  $S$  are equivalent to Sinkov's generators\* for  $(2, 3, 2k; 3)$ .

2.9. **An extension of the criterion.** Since  $(l, m, n; q)$  involves  $l, m, n$  symmetrically, Theorem D suggests that possibly  $(l, m, n; q)$  is finite (or collapses) whenever

$$(2.91) \quad \cos 2\pi/l + \cos 2\pi/m + \cos 2\pi/n \\ + (1 + 4 \cos \pi/l \cos \pi/m \cos \pi/n) \cos \pi/q < 0. \dagger$$

This inequality, with  $l=2$ , is just (2.61). The solutions with  $2 < l \leq m \leq n$  and  $q > 1$  are

$$(3, 3, 3; q), \quad (3, 3, n; 2), \quad (3, 3, 4; 3), \quad (3, 4, 4; 2), \quad (3, 4, 5; 2).$$

The first and third of these have already been described.

Sinkov has shown‡ that  $(3, 3, n; 2)$  collapses unless  $n=2, 3, 6$ , or 12. The relations

$$R^3 = S^3 = (R^{-1}S^{-1}RS)^2 = 1$$

suffice to define  $(3, 3, 12; 2)$ , of order 288. As a rotation group in four dimensions§ this is denoted by  $[3, 4, 3]''$ . It is a subgroup of index four in  $(4, 8 | 2, 3)$ . Its central quotient group  $(3, 3, 6; 2)$  is the direct product of two tetrahedral groups.|| By applying the method of §2.7 to the group

$$R^3 = S^3 = T^3 = (R^{-1}S)^2 = (S^{-1}T)^2 = (T^{-1}R)^2 = 1,$$

we find

$$\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}'$$

or  $[3^{1,1,1}]'$  (of order 96)¶ to be an invariant subgroup in  $[3, 4, 3]''$ .

\* Sinkov [2], p. 82.

† The term  $4 \cos \pi/l \cos \pi/m \cos \pi/n \cos \pi/q$  was added on June 27, 1938. Without it the criterion would admit  $(3, 3, 3; \infty)$  and  $(3, 3, 4; 4)$ . A proof that the latter group is infinite will be published elsewhere. The proviso  $q > 1$  has to be inserted because the criterion would admit  $(l, \infty, \infty; 1)$ .

‡ Sinkov [2], pp. 76-78.

§ Coxeter [5], pp. 68-70. The subgroup

$$R^{12} = S^{12} = T^{12} = RST = (R^{-1}S)^2 = (S^{-1}T)^2 = (T^{-1}R)^2 = 1,$$

of order 96, is the rotation group whose central quotient group is no. XIV ( $m=2$ ) of Goursat [1], p. 65.

|| Goursat's no. XX.

¶ Coxeter [1], p. 149, (16.75),  $n=p=q=1$ ,  $h_0=i_0=j_0=3$ . This is the subgroup  $D$  of Sinkov [2], p. 76.

The relations

$$R^4 = S^4 = (RS)^3 = (R^{-1}S^{-1}RS)^2 = 1,$$

which define  $(3, 4, 4; 2)$ , are satisfied by the permutations

$$R = (1\ 2\ 3\ 4)(7\ 8\ 9\ 10)(6\ 12),$$

$$S = (4\ 5\ 6\ 7)(3\ 8\ 11\ 12)(1\ 9),$$

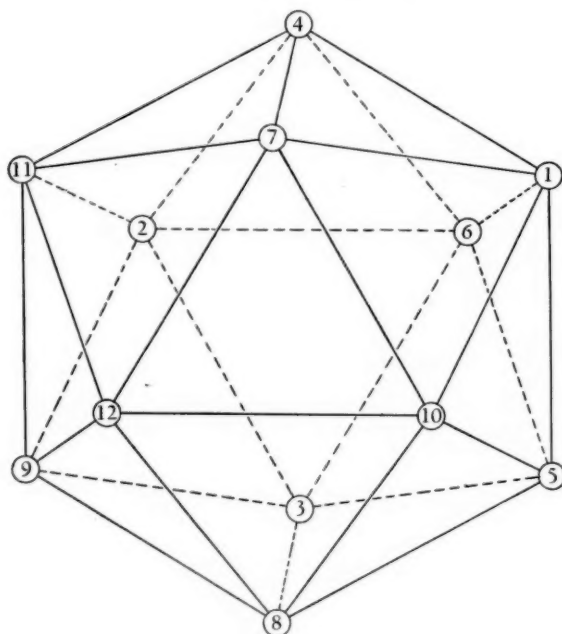


FIG. 1  
The icosahedron

which may be regarded as operating on the vertices of an icosahedron, as in Fig. 1. Although  $R$  and  $S$  are not themselves symmetries of the icosahedron, the combinations

$$RS = (1\ 2\ 8)(3\ 5\ 6)(4\ 9\ 10)(7\ 11\ 12),$$

$$S^{-1}R = (1\ 10\ 7\ 12\ 11\ 9\ 2\ 3\ 6\ 5)(4\ 8)$$

are, respectively, a rotation and a rotary reflection; these generate the extended icosahedral group  $[3, 5]$ , which is the direct product of the group of order two generated by

$$I = (R^{-1}S)^5 = (S^{-1}R)^5 = (1\ 9)(2\ 10)(3\ 7)(4\ 8)(5\ 11)(6\ 12)$$

and the icosahedral group generated by  $RS$  and

$$(R^{-1}S)^4 = S^{-1}RI.$$

Thus the order of  $(3, 4, 4; 2)$  is at least 240. The fact that it is exactly 240 may be established by enumerating the twenty cosets of the tetrahedral subgroup generated by  $R^{-1}S^{-1}$  and  $RS$ .

The operator  $I$  interchanges pairs of opposite vertices of the icosahedron, and generates the central, of order two. The quotient group,\* of order 120, is generated by the permutations

$$(1\ 2\ 3\ 4), \quad (3\ 4\ 5\ 6).$$

It is  $G_{5!}$ , since the relations

$$R^4 = S^4 = (RS)^3 = (R^{-1}S^{-1}RS)^2 = (R^{-1}S)^5 = 1$$

are also satisfied by

$$(1\ 2\ 3\ 4), \quad (5\ 4\ 3\ 2).$$

To sum up, the group  $(3, 4, 4; 2)$ , of order 240, has  $G_{5!/2} \times G_2$  as a subgroup, and has a central of order two whose quotient group is  $G_{5!}$ ; but it is not  $G_{5!} \times G_2$  (since  $R^2S$  is of period twelve).

Finally,  $(3, 4, 5; 2)$  is isomorphic with the alternating group of degree six, as generated by

$$(2.92) \quad R = (1\ 4)(2\ 6\ 3\ 5), \quad S = (1\ 2\ 3\ 4\ 5).$$

By enumerating the thirty cosets of the tetrahedral subgroup generated by  $R^{-1}S^{-1}$  and  $RS$ , we easily find that the order is just 360. Hence

$$(2.93) \quad (3, 4, 5; 2) \sim G_{6!/2}.$$

Theorem D can thus be extended as follows:

THEOREM D'. For all infinite groups  $(l, m, n; q)$  ( $q > 1$ ),

$$\cos 2\pi/l + \cos 2\pi/m + \cos 2\pi/n + (1 + 4 \cos \pi/l \cos \pi/m \cos \pi/n) \cos \pi/q \geq 0.$$

We observe that this expression vanishes for the infinite groups  $(3, 3, 3; \infty)$ ,  $(4, 4, 4; 2)$ .

All the known finite groups  $(l, m, n; q)$  are collected together in Table II, at the end of the paper.

---

\*  $(4, 4 | 3, 5; 2)$ .

CHAPTER III.  $G^{m,n,p}$ 

3.1. The derivation of  $G^{m,n,p}$  ( $p$  even) from its subgroup  $(2, m, n; p/2)$ .  $G^{m,n,p}$  means the group defined by

$$(3.11) \quad A^m = B^n = C^p = (AB)^2 = (BC)^2 = (CA)^2 = (ABC)^2 = 1.$$

This is symmetrical between  $m, n, p$ : for cyclic permutation, obviously; and for transposition, by changing  $A, B, C$  into  $C^{-1}, B^{-1}, A^{-1}$ , respectively. Since

$$ABCABC = B^{-1}A^{-1} \cdot A^{-1}C^{-1} \cdot C^{-1}B^{-1} = (BC^2A^2B)^{-1},$$

the defining relations imply

$$(3.12) \quad A^2B^2C^2 = 1.$$

Hence, if any one of  $m, n, p$  is odd, the corresponding one of  $A, B, C$  is expressible in terms of the other two. Thus  $G^{m,n,p}$  ( $p$  odd) is a factor group of  $(2, m, n; p)$ .

If  $p$  is even, all the relations (3.11) involve  $C$  an even number of times; hence the set of all operators which involve  $C$  an even number of times is a subgroup of index two. Since

$$CA = A^{-1}C^{-1}, \quad C^{-1}A^{-1} = AC, \quad CB = B^{-1}C^{-1}, \quad C^{-1}B^{-1} = BC, \quad C^2 = B^{-2}A^{-2},$$

this subgroup is generated by  $A$  and  $B$ , which satisfy

$$(3.13) \quad A^m = B^n = (AB)^2 = (A^2B^2)^{p/2} = 1.$$

Actually, these last relations completely define the subgroup. This fact will emerge from the following more general investigation.

The group  $(l, m, n; q)$ , in the form

$$S^m = T^n = (ST)^l = (S^{-1}T^{-1}ST)^q = 1,$$

clearly possesses an automorphism which replaces the generators by their inverses. If we adjoin an involutory operator  $R_2$  which transforms the group according to this automorphism, we obtain a larger group, say  $((l, m, n; 2q))$ , defined by

$$(3.14) \quad R_2^2 = S^m = T^n = (R_2S)^2 = (R_2T)^2 = (ST)^l = (R_2ST)^{2q} = 1.$$

To see that this involves  $l, m, n$  symmetrically, we define

$$R_1 = SR_2, \quad R_3 = R_2T.$$

Eliminating  $S$  and  $T$ , we obtain

$$(3.15) \quad R_1^2 = R_2^2 = R_3^2 = (R_3R_1)^l = (R_1R_2)^m = (R_2R_3)^n = (R_1R_2R_3)^{2q} = 1.$$

This group, then, contains  $(l, m, n; q)$  as a subgroup of index two. Writing the latter in the form (2.111), we see that

$$R = R_3R_1, \quad S = R_1R_2, \quad T = R_2R_3.$$

Thus each of  $R_1, R_2, R_3$  transforms two of  $R, S, T$  into their inverses.

By defining  $U = R_3R_2R_1$ , we may write (3.15) in the form

$$(3.16) \quad S^m = T^n = U^{2q} = (ST)^l = (TU)^2 = (US)^2 = (STU)^2 = 1.$$

Putting  $l=2$ , we have

$$((2, m, n; 2q)) \sim G^{m, n, 2q}.$$

Thus  $(2, m, n; q)$  is a subgroup of index two in  $G^{m, n, 2q}$ , and the latter can be derived from the former by adjoining an involutory operator which transforms the (two) generators into their inverses.

When  $(2, m, n; q)$  is given in the form

$$R^m = S^n = (RS)^2 = (R^2S^2)^q = 1,$$

we naturally write  $A=R, B=S$ ; and the involutory operator which transforms these into their inverses is  $BCA$ . (For  $BCA \cdot ABCA = BCABC = A^{-1}$ , and  $BCAB \cdot BCA = CABCA = B^{-1}$ .) On the other hand, when the same group is given in the form

$$S^n = T^2 = (ST)^m = (S^{-1}TST)^q = 1,$$

we write  $A=TS^{-1}, B=S$ ; and the involutory operator which transforms  $S$  into its inverse, while leaving  $T (=AB)$  unchanged, is  $CA$ .

**3.2. The extended polyhedral groups.** The extended polyhedral group  $[m, n]$  or\*



is defined by

$$(3.21) \quad R_1^2 = R_2^2 = R_3^2 = (R_1R_2)^m = (R_2R_3)^n = (R_3R_1)^2 = 1.$$

If we write

$$A = R_1R_2, \quad B = R_2R_3, \quad C = R_3R_1,$$

so that

$$R_1 = BC, \quad R_2 = CA, \quad R_3 = AB,$$

this definition becomes

$$(3.22) \quad A^m = B^n = (AB)^2 = (BC)^2 = (CA)^2 = (ABC)^2 = 1.$$

\* Coxeter [2], pp. 589, 619.

Hence

$$[m, n] \sim G^{m, n, p},$$

where  $p$ , the period of  $C$ , remains to be determined.

This is a "group of genus zero."\* If we force  $p$  to have a smaller value than its "natural" value, we obtain the general group  $G^{m, n, p}$ , which is a factor group of  $[m, n]$ . When  $1/m + 1/n \leq 1/2$ , the "natural" value is infinite, and

$$[m, n] \sim G^{m, n, \infty}.$$

For the present, however, we concentrate our attention on the case when  $1/m + 1/n > 1/2$ .

Since each of the relations (3.21) involves an even number of  $R$ 's, the period of  $R_3 R_2 R_1$ , deduced as a consequence of those relations, must be even. The ordinary (unextended) polyhedral group  $[m, n]'$  is generated by the rotations  $A$  and  $B$ ; hence

$$[m, n]' \sim (2, m, n; p/2).$$

Using the values of  $q$  ( $q = p/2$ ) that were found in §2.3, we see that the (finite) extended polyhedral groups are

$$G^{2, n, 2n} \text{ (} n \text{ odd)}, \quad G^{2, n, n} \text{ (} n \text{ even)}, \\ G^{3, 3, 4}, \quad G^{3, 4, 6}, \quad G^{3, 5, 10}.$$

The last four of these are covered by the formula

$$(3.23) \quad \cos 2\pi/p = 1 + \cos 2\pi/m + \cos 2\pi/n.$$

In §4.7, we shall see that this formula has a geometrical significance, in spite of its failure when  $m=2$  and  $n$  is odd.

Since  $G^{2, n, p}$  has a subgroup  $(2, n, p; 1)$ , (2.22) and (2.23) show that the extended dihedral groups

$$(3.24) \quad \left\{ \begin{array}{l} G^{2, n, 2n} \text{ (} n \text{ odd)} \\ G^{2, n, n} \text{ (} n \text{ even)} \end{array} \right\} \sim [n] \times G_2$$

are the *only* groups of this form; any other values for  $p$  (given  $n$ ) would cause collapse. It is clear, also, that  $G^{3, 3, p}$ ,  $G^{3, 4, p}$ ,  $G^{3, 5, p}$  must collapse whenever  $p$  is not a divisor of 4, 6, 10, respectively; and we have just seen that they collapse when  $p=2$ . Apart from the extended polyhedral groups themselves, there remains only  $G^{3, 5, 5}$ . This, being a proper factor group of the extended

\* Dyck [1], p. 34.

icosahedral group  $G^{3,5,10}$ , must be the icosahedral group itself.

The following generating permutations are easily verified:

	$G^{3,3,4} \sim G_4$	$G^{3,4,6} \sim G_4 \times G_2$	$G^{3,5,5} \sim G_{5/2}$	$G^{3,5,10} \sim G_{5/2} \times G_2$
<i>A</i>	(3 2 1)	(4 3 2)	(5 3 2)	(5 3 2)
<i>B</i>	(4 3 2)	(1 2 3 4)	(1 2 3 4 5)	(1 2 3 4 5)
<i>C</i>	(1 2 3 4)	(3 2 1)(5 6)	(1 2 4 3 5)	(1 2 4 3 5)(6 7)

**3.3. Cases in which  $m, n$  are odd, while  $p$  is even.** By (2.53), the group  $(2, 5, 5; 2)$ , of order 80, is generated (in the form (3.13)) by

$$A = (0\ 1\ 2\ 3\ 4)(5\ 6\ 7\ 8\ 9), \quad B = (0\ 9\ 8\ 7\ 6)(5\ 4\ 3\ 2\ 1).$$

These are transformed into their inverses by

$$(0\ 5)(1\ 9)(2\ 8)(3\ 7)(4\ 6),$$

which, being an odd permutation,\* may be identified with  $BCA$ . Since  $B$  is of odd period, the whole group  $G^{3,5,4}$  (or  $G^{4,5,5}$ ) is generated by  $A$  and

$$C = (0\ 3\ 5\ 8)(1\ 7)(2\ 6).$$

But these permutations satisfy

$$C^4 = A^5 = (CA)^2 = (C^{-1}A)^4 = 1.$$

Hence

$$(3.31) \quad G^{4,5,5} \sim (4, 5 \mid 2, 4).$$

This may be compared with the relation

$$G^{4,3,3} \sim (4, 3 \mid 2, 4),$$

which expresses the well known simple isomorphism between the extended tetrahedral and unextended octahedral groups.

In other cases we suppose  $(2, m, n; q)$  to be given in the form

$$S^n = T^2 = (ST)^m = (S^{-1}TST)^q = 1,$$

where  $S=B$ ,  $T=AB$ , and we derive  $G^{m,n,2q}$  by adjoining the operator  $CA$  which transforms  $S$  into its inverse, leaving  $T$  unchanged. We use a tabular arrangement, as in §2.5. ( $G^{3,3,4}$  and  $G^{3,5,10}$  have already been described; the latter is included now for the sake of comparison.)

\* The even permutation  $(1\ 4)(2\ 3)(6\ 9)(7\ 8)$  could have been used just as well, but it is less obviously outside the group  $(2, 5, 5; 2)$ .

$(2, m, n; q)$	$(2, 3, 5; 5) \sim G_{51/2}$	$(2, 3, 7; 4) \sim \mathfrak{P}_1(7)$
$S$ $T$	$(1\ 2\ 3\ 4\ 5)$ $(1\ 2)(4\ 5)$	$(0\ 1\ 2\ 3\ 4\ 5\ 6)$ $(0\ \infty)(1\ 6)(2\ 3)(4\ 5)$
Automorphism ( $S, S^{-1}$ )	$(1\ 5)(2\ 4)$ (inner)	$(1\ 6)(2\ 5)(3\ 4)$ (outer)
$G^{m,n,2q}$	$G^{3,5,10} \sim G_{51/2} \times G_2$	$G^{3,7,8} \sim \tilde{\mathfrak{P}}_1(7)$

$(2, m, n; q)$	$(2, 3, 7; 6) \sim \mathfrak{P}_1(13)$	$(2, 3, 7; 7) \sim \mathfrak{P}_1(13)$
$S$ $T$	$(0\ 10\ 2\ 9\ 3\ 8\ 5)(1\ 7\ 11\ 4\ 6\ \infty\ 12)$ $(1\ 12)(2\ 11)(3\ 10)(4\ 9)(5\ 8)(6\ 7)$	$(0\ 9\ 11\ 5\ 12\ 1\ 10)(2\ 7\ 6\ 4\ 3\ 8\ \infty)$ $(1\ 12)(2\ 11)(3\ 10)(4\ 9)(5\ 8)(6\ 7)$
Automorphism ( $S, S^{-1}$ )	$(0\ \infty)(6\ 10)(2\ 4)(9\ 11)(3\ 7)(1\ 8)(5\ 12)$ (outer)	$(0\ \infty)(8\ 9)(3\ 11)(4\ 5)(6\ 12)(1\ 7)(2\ 10)$ (outer)
$G^{m,n,2q}$	$G^{3,7,12} \sim \tilde{\mathfrak{P}}_1(13)$	$G^{3,7,14} \sim \tilde{\mathfrak{P}}_1(13)$

$(2, m, n; q)$	$(2, 5, 9; 2) \sim \mathfrak{P}_1(19)$
$S$ $T$	$(1\ 16\ 9\ 11\ 5\ 4\ 7\ 17\ 6)(2\ 13\ 18\ 3\ 10\ 8\ 14\ 15\ 12)$ $(0\ 7)(1\ \infty)(2\ 14)(3\ 17)(4\ 18)(5\ 9)(6\ 15)(8\ 11)(10\ 13)(12\ 16)$
Automorphism ( $S, S^{-1}$ )	$(0\ \infty)(1\ 7)(4\ 16)(5\ 9)(6\ 17)(3\ 15)(10\ 14)(12\ 18)(2\ 13)$ (outer)
$G^{m,n,2q}$	$G^{4,5,9} \sim \tilde{\mathfrak{P}}_1(19)$

$(2, m, n; q)$	$(2, 3, 11; 4) \sim \mathfrak{P}_1(23)$
$S$ $T$	$(1\ 9\ 12\ 16\ 6\ 8\ 3\ 4\ 13\ 2\ 18)(5\ 22\ 14\ 11\ 7\ 17\ 15\ 20\ 19\ 10\ 21)$ $(0\ 3)(1\ \infty)(2\ 22)(4\ 8)(5\ 12)(6\ 19)(7\ 16)(9\ 18)(10\ 11)(13\ 20)(14\ 15)(17\ 21)$
Automorphism ( $S, S^{-1}$ )	$(0\ \infty)(1\ 3)(4\ 18)(2\ 13)(8\ 9)(6\ 12)(5\ 19)(10\ 21)(20\ 22)(14\ 15)(11\ 17)$ (outer)
$G^{m,n,2q}$	$G^{3,8,11} \sim \tilde{\mathfrak{P}}_1(23)$

The generating permutations are taken from (2.39), (2.391), (2.64), (2.66). In each case they are even permutations; so whenever the desired automor-

phism is effected by an odd permutation, we can be sure that it is an outer automorphism. (This permutation is easily found by observing which symbols are left unchanged by one of the generators.)

The apparently possible groups

$$G^{3,6,n} \ (n \text{ odd}), \ G^{3,7,10}, \ G^{3,8,9}, \ G^{4,4,n} \ (n \text{ odd}), \ G^{4,5,7}$$

collapse with their would-be subgroups (see §2.6, §2.3, and the Appendix)

$$(2, 3, n; 3), \quad (2, 3, 7; 5), \quad (2, 3, 9; 4), \quad (2, 4, n; 2), \quad (2, 5, 7; 2).$$

3.4. **Groups of genus one:**  $G^{4,4,2k}$  and  $G^{3,6,2k}$ . The general results of §2.5 and §3.1 are illustrated as "genealogies" in Fig. 2, where each single-headed arrow stands for "is a subgroup of index two in," while each double-headed arrow stands for "is a subgroup of index three in."

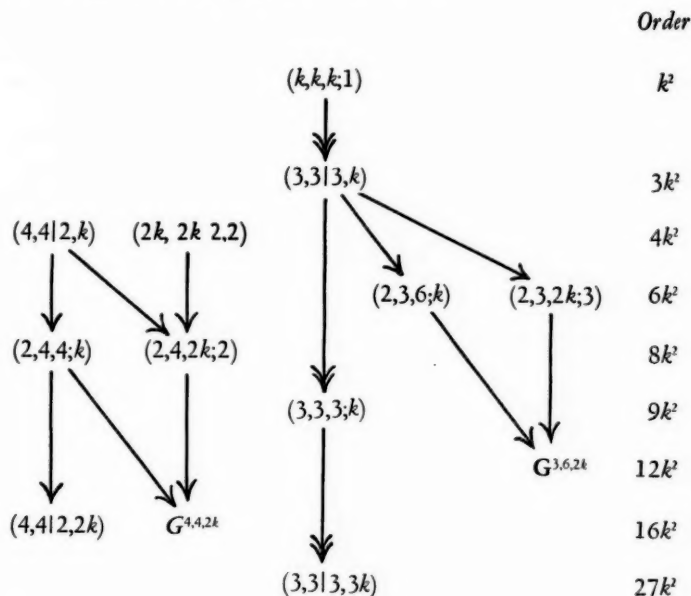


FIG. 2  
Groups of genus one

When  $k$  is odd, (2.86) and (2.861) show that  $(2, 4, 2k; 2)$  is generated by

$$AB = (1 \ 2k)(2 \ 2k - 1) \cdots (k \ k + 1),$$

$$B = (1 \ 2)(3 \ 4 \ 5 \ 6) \cdots (2k - 3 \cdots 2k).$$

$B$  is transformed into its inverse, and  $AB$  into itself, by

$$(1\ 2)(3\ 4)(5\ 6) \cdots (2k-1\ 2k).$$

Calling this  $C_1A$ , where

$$A = (1\ 2k-1)(3\ 2k-3) \cdots (k-2\ k+2)(2\ 2k-2\ 6\ 2k-6 \cdots 2k-4\ 4\ 2k),$$

we find

$$C_1 = (1\ 2k)(2\ 2k-1\ 4\ 2k-3)(3\ 2k-4\ 5\ 2k-2)(6\ 2k-5\ 8\ 2k-7) \cdots.$$

Since  $C_1^{-1}B$  is of period  $k$ ,  $C_1$  and  $B$  generate  $(4, 4 \mid 2, k)$ , whereas the  $C$  and  $B$  of  $G^{2k, 4, 4}$  generate the larger group  $(2, 4, 4; k)$ . Thus  $(2, 4, 2k; 2)$  is transformed according to an inner automorphism. Now,  $k$  being odd, the central of  $(2, 4, 2k; 2)$  is of order 1. Hence  $G^{2k, 4, 4}$  is a direct product, namely

$$(3.41) \quad G^{4, 4, 2k} \sim (2, 4, 2k; 2) \times G_2, \quad k \text{ odd.}^*$$

By (2.85) and (2.851), on the other hand, whatever be the parity of  $k$ ,  $(2, 4, 2k; 2)$  is generated by

$$B = (a_1\ a_{2k})(a_2\ a_{2k-1}) \cdots (a_k\ a_{k+1})(b_1\ b_2 \cdots b_{2k}),$$

$$AB = (a_1\ b_1)(a_2\ b_2) \cdots (a_{2k}\ b_{2k}).$$

$B$  is transformed into the inverse, and  $AB$  into itself, by

$$(a_1\ a_{2k})(a_2\ a_{2k-1}) \cdots (a_k\ a_{k+1})(b_1\ b_{2k})(b_2\ b_{2k-1}) \cdots (b_k\ b_{k+1}).$$

Calling this  $CA$ , where

$$A = (a_1\ b_{2k})(b_1\ a_{2k}\ b_{2k-1}\ a_2)(b_2\ a_{2k-1}\ b_{2k-2}\ a_3) \cdots (b_{k-1}\ a_{k+2}\ b_{k+1}\ a_k)(a_{k+1}\ b_k),$$

we find

$$C = (a_1\ b_1)(a_2\ b_2\ a_{2k}\ b_{2k})(a_3\ b_3\ a_{2k-1}\ b_{2k-1}) \cdots (a_k\ b_k\ a_{k+2}\ b_{k+2})(a_{k+1}\ b_{k+1}).$$

Since  $A^{-1}C = (a_1\ a_2 \cdots a_{2k})(b_1\ b_2 \cdots b_{2k})$ ,  $A$  and  $C$  generate the whole group  $(2, 4, 4; k)$ , and therefore  $A, B, C$  generate  $G^{4, 2k, 4}$  (that is,  $G^{4, 4, 2k}$ ).

Similarly, (2.89) and (2.891) show that  $(2, 3, 2k; 3)$  is generated by

$$B = (3k\ 1\ 2)(3\ 4\ 5) \cdots (3k-3\ 3k-2\ 3k-1),$$

$$AB = (1\ 3k-1)(2\ 3k-2)(3\ 3k-3) \cdots.$$

$B$  is transformed into its inverse, and  $AB$  into itself, by

$$(1\ 2)(4\ 5) \cdots (3k-2\ 3k-1).$$

Calling this  $CA$ , where

---

\* Cf. (2.84).

$A = (3k \ 2 \ 3k - 3 \ 5 \ 3k - 6 \ 8 \cdots 3 \ 3k - 1)(1 \ 3k - 2)(4 \ 3k - 5)(7 \ 3k - 8) \cdots$ ,  
we find

$$C = (1 \ 3k \ 3k - 1)(2 \ 3k - 2 \ 3 \ 3k - 4 \ 4 \ 3k - 3)(5 \ 3k - 5 \ 6 \ 3k - 7 \ 7 \ 3k - 6) \cdots.$$

Since  $(2, 3, 2k; 3)$  has, in general, no operator of period six,\* we can be sure that  $A, B, C$  (or just  $A$  and  $C$ , since the period of  $B$  is odd) generate  $G^{2k, 3, 6}$  (that is,  $G^{3, 6, 2k}$ ).

**3.5. The derivation of  $G^{m, 2n, 2k}$  from its subgroup  $(m, m | n, k)$ .** We saw, in §2.5, that the groups  $(2, m, 2n; k)$ ,  $(2, m, 2k; n)$  can be derived from  $(m, m | n, k)$  by adjoining involutory operators  $T, T'$ , respectively;  $T$  interchanges  $R, S$ , while  $T'$  transforms each into the inverse of the other. The product  $TT'$  (or  $T'T$ ) transforms each into its own inverse (see §1.2). Let us now adjoin these two permutable operators simultaneously, that is, adjoin the four-group which they generate.

Along with the relations

$$R^m = S^m = (RS)^n = (R^{-1}S)^k = 1,$$

which define  $(m, m | n, k)$ , we now have

$$T^2 = T'^2 = 1, \quad TT' = T'T, \quad TST = R = T'S^{-1}T'.$$

Eliminating  $R$ , we obtain

$$S^m = T^2 = T'^2 = (ST)^{2n} = (ST')^{2k} = (TT')^2 = (STT')^2 = 1.$$

We can identify this with  $G^{m, 2n, 2k}$  by writing

$$(3.51) \quad \begin{aligned} S &= A^{-1}, & T &= AB, & T' &= CA; \\ A &= S^{-1}, & B &= ST (= TR), & C &= T'S (= R^{-1}T'). \end{aligned}$$

Thus  $(m, m | n, k)$  is an invariant subgroup of index four in  $G^{m, 2n, 2k}$ .

When

$$(2, m, 2n; k) \sim (2, m, 2k; n)$$

(in particular, when  $n=k$ ), it may happen that  $TT'$  transforms  $(m, m | n, k)$  according to an inner automorphism. Then  $T'$  transforms  $(2, m, 2n; k)$  according to an inner automorphism, and, if the order of the central of  $(m, m | n, k)$  is odd,

$$(3.52) \quad G^{m, 2n, 2k} \sim (2, m, 2n; k) \times G_2.$$

Applying this principle to  $(5, 5 | 2, 3)$ ,  $(3, 3 | 4, 4)$ ,  $(3, 3 | 4, 5)$  (see (2.56)), we obtain

\* See Coxeter [5], p. 67, where such an operator would appear as a hexagonal rotation.

$$(3.53) \quad G^{4,5,6} \sim G_5! \times G_2,$$

$$(3.54) \quad G^{3,8,8} \sim \tilde{\mathfrak{P}}_1(7) \times G_2,$$

$$(3.55) \quad G^{3,8,10} \sim (2, 3, 8; 5) \times G_2.$$

When the groups  $(2, m, 2n; k)$ ,  $(2, m, 2k; n)$  are distinct, it may happen that  $T$  and  $T'$  transform  $(m, m | n, k)$  according to outer and inner automorphisms, respectively. Then  $T'$  transforms  $(2, m, 2n; k)$  according to an inner automorphism, and (with the usual proviso about the central) (3.52) follows again.

Applying this principle to  $(5, 5 | 4, 2)^*$  and  $(7, 7 | 2, 3)$ , we obtain

$$(3.56) \quad G^{4,5,8} \sim (2, 5, 8; 2) \times G_2,$$

$$(3.57) \quad G^{4,6,7} \sim \tilde{\mathfrak{P}}_1(13) \times G_2.$$

The same principle, applied to  $(4, 4 | k, 2)$ , provides an alternative proof for (3.41).

**3.6. Definitions for  $G^{m,n,p}$  ( $m$  odd) in terms of two generators.** When  $m$  is odd, the relation (3.12) gives

$$A = A^{-(m-1)} = (B^2C^2)^{(m-1)/2}.$$

Thus  $G^{m,n,p}$  is generated by  $B$  and  $C$ . By direct substitution, (3.11) becomes

$$\begin{aligned} B^n &= C^p = (BC)^2 = (B^2C^2)^m = [(B^2C^2)^{(m-1)/2}B]^2 \\ &= [C(B^2C^2)^{(m-1)/2}]^2 = 1, \\ [C(B^2C^2)^{(m-1)/2}B]^2 &= 1. \end{aligned}$$

The last relation is superfluous, since  $(BC)^2 = 1$  implies

$$\begin{aligned} C(B^2C^2)^{(m-1)/2}B &= CB \cdot BC \cdot \dots \cdot BC \cdot CB \\ &= B^{-1}C^{-1} \cdot C^{-1}B^{-1} \cdot \dots \cdot C^{-1}B^{-1} \cdot B^{-1}C^{-1} \\ &= B^{-1}(C^{-2}B^{-2})^{(m-1)/2}C^{-1}. \end{aligned}$$

We may also omit any one of the three relations

$$(B^2C^2)^m = 1, \quad [(B^2C^2)^{(m-1)/2}B]^2 = 1, \quad [C(B^2C^2)^{(m-1)/2}]^2 = 1,$$

since  $(BC)^2 = 1$  implies

$$\begin{aligned} [(B^2C^2)^{(m-1)/2}B]^2 [C(B^2C^2)^{(m-1)/2}]^2 &= (B^2C^2)^{(m-1)/2}B \cdot BC \cdot C(B^2C^2)^{(m-1)/2} \\ &= (B^2C^2)^m. \end{aligned}$$

Thus  $G^{m,n,p}$  ( $m$  odd) is defined by

$$(3.61) \quad B^n = C^p = (BC)^2 = (B^2C^2)^m = [B(B^2C^2)^{(m-1)/2}]^2 = 1$$

\* Or to  $(5, 5 | 2, 4)$ , interchanging the roles of  $T$  and  $T'$ .

or

$$(3.611) \quad B^n = C^p = (BC)^2 = [B(B^2C^2)^{(m-1)/2}]^2 = [(B^2C^2)^{(m-1)/2}C]^2 = 1.$$

Of these two definitions, the latter is more symmetrical (especially when  $n=p$ ); but the former is simpler, and exhibits  $G^{m,n,p}$  as a factor group of  $(2, n, p; m)$ . In particular, putting  $m=3$ , we see that  $G^{3,n,p}$  is defined by

$$(3.62) \quad B^n = C^p = (BC)^2 = (B^2C^2)^3 = (B^3C^2)^2 = 1$$

or by

$$(3.621) \quad B^n = C^p = (BC)^2 = (B^3C^2)^2 = (B^2C^3)^2 = 1.$$

In order to conform with Sinkov's notation,\* let us write  $B=P$ ,  $C^2=Q$ . When  $p$  is even, we obtain the subgroup  $(2, m, n; p/2)$  in the form

$$(3.63) \quad P^n = Q^{p/2} = (P^2Q)^m = [P(P^2Q)^{(m-1)/2}]^2 = 1.$$

To see that this definition is sufficient, we put  $(P^2Q)^{(m-1)/2}=S$ , so that  $Q=P^{-2}S^{-2}$ , and deduce

$$S^m = P^n = (SP)^2 = (S^2P^2)^{p/2} = 1.$$

When  $p$  (as well as  $m$ ) is odd, so that  $C=C^{p+1}=Q^{(p+1)/2}$ , (3.61) becomes

$$(3.64) \quad P^n = Q^p = (P^2Q)^m = (PQ^{(p+1)/2})^2 = [P(P^2Q)^{(m-1)/2}]^2 = 1.$$

This exhibits  $G^{m,n,p}$  as a factor group of  $(2, m, n; p)$  (see (3.63)), with the very simple extra relation

$$(PQ^{(p+1)/2})^2 = 1.$$

Putting  $m=3$ , we have†

$$(3.65) \quad P^n = Q^p = (P^2Q)^3 = (P^3Q)^2 = 1$$

for  $(2, 3, n; p)$ , and

$$(3.66) \quad P^n = Q^p = (P^2Q)^3 = (P^3Q)^2 = (PQ^{(p+1)/2})^2 = 1$$

for  $G^{3,n,p}$  ( $p$  odd).

In (3.66), any one of the three expressions

$$(P^2Q)^3, \quad (P^3Q)^2, \quad (PQ^{(p+1)/2})^2$$

may be replaced by  $(P^2Q^{(p+3)/2})^2$ ; the first two, because

$$(B^2C^2B)^2(CB^2C^2)^2 = (B^2C^2)^3$$

\* Sinkov [3], p. 68. In " $Q=C^2, P=A$ ,"  $A$  is a misprint for  $B$ .

† For the case when  $n=7$ , see Brahana [2], p. 349. Our  $P$  (or  $B$ ) is his  $R$ .

(as we saw above, before putting  $m=3$ ), and the last by the following argument, due to Sinkov. He has shown\* that the relations  $(P^2Q)^3 = (P^3Q)^2 = 1$  imply  $(PQP^2Q^\alpha)^2 = 1$  (for any  $\alpha$ ). If also  $(P^2Q^\alpha)^2 = 1$ , where  $\alpha = (p+3)/2$ , we have

$$(PQ^{(p+1)/2})^2 = (P^{-1} \cdot P^2Q^\alpha \cdot Q^{-1})^2 = (P^{-1}Q^{-\alpha}P^{-2}Q^{-1})^2 = (QP^2Q^\alpha P)^{-2} = 1.$$

Thus  $G^{3,n,p}$  ( $p$  odd) is defined by

$$(3.661) \quad P^n = Q^p = (P^2Q)^3 = (P^3Q)^2 = (P^2Q^{(p+3)/2})^2 = 1.$$

This is the definition used by Sinkov in his proof† that  $G^{3,7,7}$  and  $G^{3,7,11}$  collapse, and in his proof‡ that

$$(3.67) \quad G^{3,7,9} \sim \mathfrak{P}_1(8).$$

(This is the simple group of order 504.)

Moreover, Sinkov has achieved a further simplification of the definition, in the case when  $n$  (as well as  $p$ ) is odd; namely, he finds§ that the period of  $Q$  need not be specified. Thus  $G^{3,n,p}$  ( $n, p$  odd) is defined by

$$(3.68) \quad P^n = (P^2Q)^3 = (P^3Q)^2 = (PQ^{(p+1)/2})^2 = 1.$$

For example,  $\mathfrak{P}_1(8)$  is defined by

$$(3.69) \quad P^7 = (P^2Q)^3 = (P^3Q)^2 = (PQ^5)^2 = 1$$

or, interchanging  $n$  and  $p$ , by  $P^9 = (P^2Q)^3 = (P^3Q)^2 = (PQ^4)^2 = 1$ .

**3.7. Cases in which  $m, n, p$  are all odd:**  $G^{3,5,5}$ ,  $G^{3,7,9}$ ,  $G^{3,7,13}$ ,  $G^{3,7,15}$ ,  $G^{3,9,9}$ ,  $G^{3,5,5}$ . In §3.2, we found  $G^{3,5,5}$  to be the icosahedral group, as generated by

$$A = (2 \ 5 \ 3), \quad B = (1 \ 2 \ 3 \ 4 \ 5), \quad C = (1 \ 2 \ 4 \ 3 \ 5).$$

Clearly,  $A$  and  $B$  generate the same group in the form

$$A^3 = B^5 = (AB)^2 = 1,$$

while  $B$  and  $C$  generate it *qua*  $(5, 5 \mid 2, 3)$ .

We easily verify that  $\mathfrak{P}_1(8)$ , in the form  $G^{3,7,9}$ , is generated by the linear fractional substitutions

$$A = \left( \frac{1}{x+1} \right), \quad B = (\alpha^2 x + 1), \quad C = \left( \frac{\alpha^5 x + 1}{\alpha^3 x + \alpha} \right),$$

\* Sinkov [3], p. 68.

† Sinkov [5].

‡ Sinkov [3], p. 70.

§ Sinkov [3], p. 69.

where  $\alpha$  is a primitive root in the Galois field  $GF(2^3)$ , defined by

$$\alpha^3 + \alpha + 1 \equiv 0 \pmod{2}.$$

In the form (3.69), we therefore have

$$P = (\alpha^2 x + 1), \quad Q = \left( \frac{\alpha}{\alpha^4 x + 1} \right).$$

We still more easily verify that the defining relations for  $G^{3,7,13}$  are satisfied by the linear fractional substitutions

$$A = \left( \frac{1}{1-x} \right), \quad B = \left( \frac{1}{x} - 1 \right), \quad C = (x+1) \pmod{13}.$$

Hence  $\mathfrak{P}_1(13)$  occurs as a factor group. (In Theorem G, below, we shall generalize this result.) Sinkov has proved\* that the order of  $G^{3,7,13}$  is 2184, which is only twice that of  $\mathfrak{P}_1(13)$ . In fact,  $\mathfrak{P}_1(13)$  is the quotient group of the central (of order two) generated by  $(P^6 Q^2)^3$ ,  $(PQ^3)^3$ ,  $(BC^6)^3$ , or  $(CBA)^7$ .

Sinkov has remarked that  $G^{3,7,13}$  does not contain  $\mathfrak{P}_1(13)$  as a subgroup. For, such a subgroup, being of index two, would be invariant; and, having an abelian quotient group (of order two), it would contain the commutator subgroup of  $G^{3,7,13}$ ; but  $G^{3,7,13}$  is known to be *perfect*,† since it is generated by two operators ( $AB$  and  $A$ ) of periods two and three, whose product is of period seven.

Thus  $G^{3,7,13}$  is not isomorphic with  $G^{3,7,12}$  or  $G^{3,7,14}$ , although it has the same order (2184). Nor is it the group of unimodular substitutions on two variables (mod 13),‡ since, if we represent  $A, B, C$  by the matrices

$$\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad \begin{pmatrix} -5 & 5 \\ 5 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

we find that

$$(CA)^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Sinkov has observed that the defining relations for  $G^{3,7,15}$  are satisfied by the linear fractional substitutions

$$A = \left( \frac{4x+16}{x+26} \right), \quad B = \left( \frac{22x+11}{12x+10} \right), \quad C = \left( \frac{10x+1}{12x+10} \right) \pmod{29}.$$

\* Sinkov [3], p. 73.

† Brahana [2], p. 347.

‡ Dickson's  $SLH(2, 13)$ ; van der Waerden's  $SL(2, 13)$ . That group is defined by

$$A^3 = B^7 = C^{13} = I^2 = 1, \quad (AB)^2 = (BC)^2 = (CA)^2 = (ABC)^2 = (CBA)^7 = I.$$

He and I have investigated this group, in the form

$$P^7 = (P^2Q)^3 = (P^3Q)^2 = (PQ^8)^2 = 1.$$

By enumerating the 406 cosets of the subgroup of order thirty generated by  $Q$  and  $PQP^2$ , we established the order as being 12180. Hence

$$(3.71) \quad G^{3,7,15} \sim \mathfrak{P}_1(29).$$

Similarly, the defining relations for  $G^{3,9,9}$  are satisfied by the linear fractional substitutions

$$A = \left( \frac{6x+13}{2x+14} \right), \quad B = \left( \frac{4x+7}{7x+3} \right), \quad C = \left( \frac{2x}{10} \right) \pmod{19}.$$

By enumerating the 190 cosets of the subgroup of order eighteen generated by  $AB$  and  $C$ , J. M. Kingston\* obtained the order 3420. Hence

$$(3.72) \quad G^{3,9,9} \sim \mathfrak{P}_1(19).$$

Brahana has shown† that  $\mathfrak{P}_1(11)$  is generated by the permutations

$$S = (a g e b k)(c j i d f), \quad T = (a c)(b d)(e g)(f h).$$

Writing

$$B = S^{-1} = (a k b e g)(j c f d i), \quad C = ST = (b k c j i)(f a e d h),$$

we observe that these permutations satisfy the relations

$$B^5 = C^5 = (BC)^2 = (B^2C^2)^5 = (B^3C^2B^2C^2)^2 = 1,$$

which, by (3.61), define  $G^{5,5,5}$ . By enumerating the 66 cosets of the subgroup of order ten generated by  $BC$  and  $CB$ , we easily find that these relations suffice to define a group of order 660. Hence‡

$$(3.73) \quad G^{5,5,5} \sim \mathfrak{P}_1(11).$$

The third generator for  $G^{5,5,5}$  (in the form (3.11)) is

$$A = (B^2C^2)^2 = (S^{-1}TST)^2 = (c k a f h)(e b j d g).$$

$A, B, C$  are cyclically permuted by  $D^{-1} = (a, b, c)(e, j, f)(g, i, h)$ . By adjoining this operator, we obtain the direct product  $\mathfrak{P}_1(11) \times G_3$  in the form

\* On July 14, 1938.

† Brahana [3], p. 546.

‡ Lewis [1], defines this group in the form  $S^{11} = T^2 = (ST)^3 = (S^4TS^8T)^2 = 1$ . Cf. Brahana [2], p. 356, where two other elegant definitions are suggested:  $P^{11} = Q^5 = (PQ)^2 = (P^2Q^2)^2 = (P^8Q)^3 = 1$  ( $P^2 = R$ ),  $R^{11} = T^2 = (RT)^3 = (R^{-1}TRT)^5 = (R^{-3}TR^3T)^3 = 1$ . For another form of this last definition, see Todd and Coxeter [1], p. 32.

$$C^5 = D^3 = (CD)^6 = (CDCD^{-1})^2 = 1. *$$

$G^{4,4,4}$  leads analogously to the group

$$C^4 = D^3 = (CD)^6 = (CDCD^{-1})^2 = 1,$$

of order 192, which is generated by the permutations

$$(a_1 b_1)(a_3 b_3)(a_2 b_2 a_4 b_4), \quad (a_2 b_2 b_3)(a_4 b_4 b_1);$$

and  $G^{2,2,2} (G^{2,2,2} \sim G_2 \times G_2 \times G_2)$  leads to the pyritohedral group (2, 3, 6; 2).

So too, the icosahedral group†

$$(3.74) \quad A^3 = B^3 = C^3 = (AB)^2 = (BC)^2 = (CA)^2 = 1$$

leads to the group

$$(3.75) \quad C^3 = D^3 = (CDCD^{-1})^2 = 1,$$

of order 180. This is  $G_{51/2} \times G_3$ , since the generators

$$A = (a_1 a_4 a_5), \quad B = (a_2 a_4 a_5), \quad C = (a_3 a_4 a_5),$$

of (3.74) are cyclically permuted by  $(a_1 a_2 a_3)$ . The relations (3.75) imply  $(CD)^{15} = 1$ ; but  $(a_3 a_4 a_5)$  and  $(a_1 a_2 a_3)$  generate the factor group

$$C^3 = D^3 = (CD)^5 = (CDCD^{-1})^2 = 1,$$

which is therefore the icosahedral group again. Clearly,  $D$  and  $(CD)^2$  satisfy the ordinary definition  $D^3 = E^3 = (DE)^2 = 1$ . More generally, the relations

$$C^k = D^3 = (CD)^5 = (CDCD^{-1})^l = 1$$

define  $(l, 5 | 3, k)$ . (The interesting cases are when  $l = 3$  and  $k = 3$  or 4.)

It is perhaps appropriate to remark here that the octahedral group  $G^{3,3,4}$  appears naturally as a factor group of the (unextended) hyper-octahedral group‡  $A^3 = B^3 = C^4 = (AB)^2 = (BC)^2 = (CA)^2 = 1$ , of order 192.

**3.8. The criterion for finiteness.** By applying §3.1 to the groups considered in Theorem C, we immediately deduce the following theorem:

**THEOREM E.** *If the smallest of  $m, n, p$  is greater than 2, while the next is greater than 3, and if these three numbers are either all even, or one even and the other two equal, the group  $G^{m,n,p}$  is finite when*

$$(3.81) \quad \cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p < 1,$$

*and infinite otherwise.*

\* In terms of  $CD$  and  $DCDCD$ , this takes the form  $S^6 = T^2 = (S^2T)^3 = (S^3T)^2 = 1$ . Hence yet another definition for  $\mathfrak{P}_1(11)$  is  $S^6 = T^2 = (ST)^{11} = (S^2T)^3 = (S^3T)^2 = 1$ .

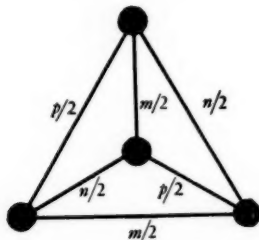
† Carmichael [1], p. 255.

‡ Coxeter [3], p. 219.

Since this result is formally more elegant than the analogous Theorems A and C, it seems worth while to give a direct proof.

When  $m, n, p$  are all even, we take the group generated by reflections

$$\begin{aligned} R_1^2 &= R_2^2 = R_3^2 = R_4^2 = (R_1 R_4)^{m/2} = (R_2 R_3)^{m/2} \\ &= (R_2 R_4)^{n/2} = (R_3 R_1)^{n/2} = (R_3 R_4)^{p/2} = (R_1 R_2)^{p/2} = 1, \end{aligned}$$



and adjoin a four-group  $M^2 = N^2 = P^2 = MNP = 1$ , such that

$$\begin{aligned} R_1 M &= M R_4, & R_2 M &= M R_3 &= A \text{ (say),} \\ R_2 N &= N R_4, & R_3 N &= N R_1 &= B \text{ (say),} \\ R_3 P &= P R_4, & R_1 P &= P R_2 &= C \text{ (say).} \end{aligned}$$

These relations imply

$$\begin{aligned} M &= BC, & N &= CA, & P &= AB, \\ R_1 &= CAB, & R_2 &= ABC, & R_3 &= BCA, & R_4 &= CA^3 B = AB^3 C = BC^3 A. \end{aligned}$$

By direct substitution, the augmented group is seen to be  $G^{m,n,p}$ .

The necessary and sufficient condition for the group generated by reflections to be finite is\*

$$\begin{vmatrix} 1 & -\cos 2\pi/p & -\cos 2\pi/n & -\cos 2\pi/m \\ -\cos 2\pi/p & 1 & -\cos 2\pi/m & -\cos 2\pi/n \\ -\cos 2\pi/n & -\cos 2\pi/m & 1 & -\cos 2\pi/p \\ -\cos 2\pi/m & -\cos 2\pi/n & -\cos 2\pi/p & 1 \end{vmatrix} > 0,$$

that is,

$$\begin{aligned} &(1 - \cos 2\pi/m - \cos 2\pi/n - \cos 2\pi/p)(1 - \cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p) \\ &\cdot (1 + \cos 2\pi/m - \cos 2\pi/n + \cos 2\pi/p)(1 + \cos 2\pi/m + \cos 2\pi/n - \cos 2\pi/p) > 0. \end{aligned}$$

\* Strictly, we should mention the necessary condition  $1/m + 1/n + 1/p > 1/2$ , which ensures that the solid angle at a vertex of the fundamental region has a spherical excess. But this condition is automatically covered by the other. (An analogous remark could have been made in §1.2.)

Since we are supposing  $m, n, p$  to be even and greater than 2, none of the cosines can be negative; so the last three factors are essentially positive, and may be discarded. (It will appear later, however, that they have a certain significance when we allow  $m, n$ , or  $p$  to be odd. See Theorem F.)

When  $m = n$ , and  $p$  is even, we observe that the subgroup  $(2, m, m; p/2)$  is also a subgroup of index two in the group

$$S^m = R_0^2 = (R_0 S^{-1} R_0 S)^2 = (R_0 S^{-2} R_0 S^2)^{p/2} = 1,$$

whose order is  $m$  times that of the group generated by reflections

$$R_i^2 = (R_i R_{i+1})^2 = (R_i R_{i+2})^{p/2} = 1, \quad R_{i+m} = R_i.$$

This is infinite when  $m > 5$ , and again when  $m = 5$  and  $p/2 > 2$ . These values agree with the trigonometrical criterion. (When  $m = 3$ , the groups collapse unless  $p = 4$ .)

This completes the proof of Theorem E, which tells us, for instance, that  $G^{4,6,6}$  and  $G^{5,5,6}$  are infinite.

The groups  $G^{4,4,2k}$  and  $G^{4,5,5}$  are the only finite groups which satisfy all the conditions of Theorem E. If we relax the conditions of parity and equality, we find that the criterion (3.81) admits the groups:

$$\begin{aligned} &G^{3,4,p} \text{ (collapsing unless } p = 3 \text{ or } 6), \\ &G^{3,5,p} \text{ (collapsing unless } p = 5 \text{ or } 10), \\ &G^{3,6,p} \text{ (collapsing if } p \text{ is odd),} \\ &G^{3,7,p} \text{ (} p \leq 12; \text{ collapsing unless } p = 8, 9, \text{ or } 12), \\ &G^{3,8,8}, G^{4,5,6}, G^{5,5,5}, \\ &G^{4,4,p} \text{ (} p \text{ odd), } G^{3,8,9}, G^{4,5,7} \text{ (collapsing),} \end{aligned}$$

all of which have already been described. Thus  $G^{m,n,p}$  is finite (or collapses) whenever (3.81) is satisfied. Moreover,\* with the exception of  $G^{2,n,2n}$  ( $n$  odd), it collapses whenever

$$(1 - \cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p)(1 + \cos 2\pi/m - \cos 2\pi/n + \cos 2\pi/p)(1 + \cos 2\pi/m + \cos 2\pi/n - \cos 2\pi/p) < 0.$$

Hence we have the following theorem:

**THEOREM F.** *For all (noncollapsing) groups  $G^{m,n,p}$ , save  $G^{2,n,2n}$  ( $n$  odd),*

$$(1 - \cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p)(1 + \cos 2\pi/m - \cos 2\pi/n + \cos 2\pi/p)(1 + \cos 2\pi/m + \cos 2\pi/n - \cos 2\pi/p) \geq 0; \dagger$$

*and for all infinite groups  $G^{m,n,p}$ ,*

\* See Figs. 3-5, below.

† The groups for which this expression vanishes, and the one exceptional group for which it is negative, are just the extended polyhedral groups  $[m, n]$  (see (3.23)).

$$(3.82) \quad \cos 2\pi/m + \cos 2\pi/n + \cos 2\pi/p \geq 1.$$

Among the infinite groups, I would mention  $G^{3,8,12}$  (see (2.67)). On the other hand, we have seen that the groups

$$G^{3,7,13}, G^{3,7,14}, G^{3,7,15}, G^{3,8,10}, G^{3,8,11}, G^{4,5,8}, G^{4,5,9}, G^{4,6,7}$$

are finite, although they satisfy (3.82).

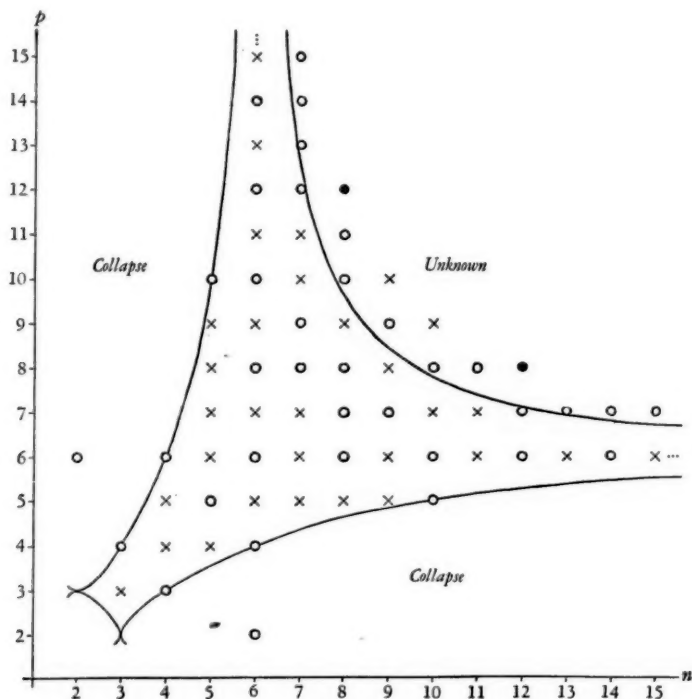


FIG. 3  
A graphical enumeration of groups  $G^{3,n,p}$

These results are collected in Table III, which shows that, with a few exceptions, the order of  $G^{m,n,p}$  increases as

$$1 - \cos 2\pi/m - \cos 2\pi/n - \cos 2\pi/p$$

diminishes.

In Figs. 3, 4, 5,\* we represent the particular groups  $G^{3,n,p}$ ,  $G^{4,n,p}$ ,  $G^{5,n,p}$  (respectively) as points with Cartesian coordinates  $(n, p)$ . The known finite groups are marked as o's, the known infinite groups as dots, and the cases of collapse as crosses (except where,  $n$  or  $p$  being small, collapse is general). The

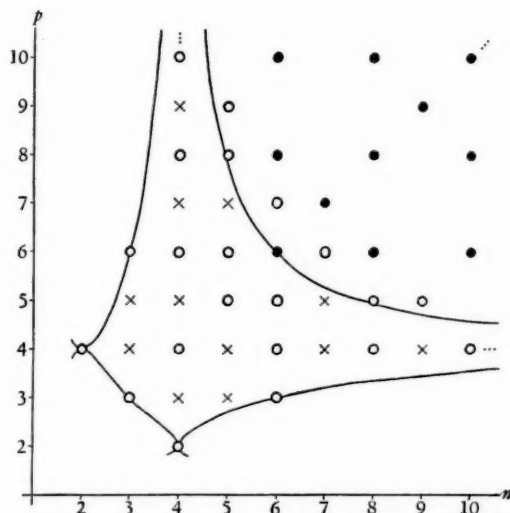


FIG. 4

A graphical enumeration of groups  $G^{4,n,p}$

trigonometrical criteria appear as curves (which roughly enclose the region in which finite groups are prevalent). Clearly, all the known finite groups  $G^{m,n,p}$  are included, some several times over.

3.9.  $G^{3,n,p}$  and the Fibonacci numbers. Let  $f_n$  denote the  $n$ th Fibonacci number, that is, the  $n$ th term of the sequence 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377,  $\dots$ , defined by the recurrence formula

$$f_1 = f_2 = 1, \quad f_n + f_{n+1} = f_{n+2}.$$

Clearly,

$$f_n = 5^{-1/2}(\tau^n - \tau'^n),$$

where  $\tau, \tau'$  are the roots  $(1 \pm 5^{1/2})/2$  of the equation

$$X^2 - X - 1 = 0.$$

\* Cf. Coxeter [7], p. 41, where the groups  $(l, m \mid 2, 3)$  and  $(l, m \mid 2, 4)$  are similarly represented.

We shall see in a moment a connection between these numbers and certain groups  $G^{3,n,p}$ , where  $p$  is an odd prime.

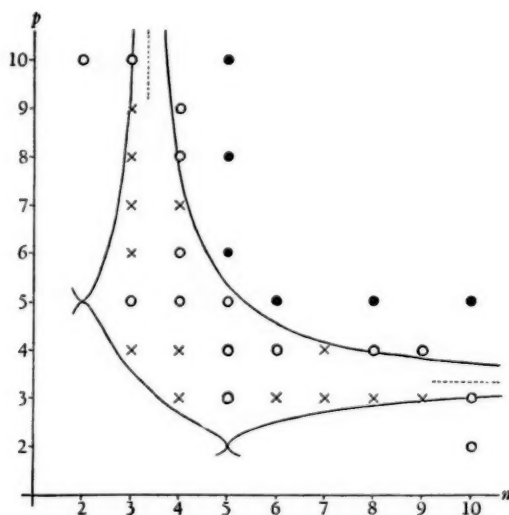


FIG. 5  
A graphical enumeration of groups  $G^{3,n,p}$

Consider the linear fractional substitutions

$$(3.91) \quad A = \left( \frac{1}{1-x} \right), \quad B = \left( \frac{1}{x} - 1 \right), \quad C = (x+1),$$

which are to be regarded as operating in the field of integers modulo  $p$ , an odd prime. It is easily verified that

$$AB = (-x), \quad BC = \left( \frac{1}{x} \right), \quad CA = \left( -\frac{1}{x} \right), \quad ABC = (1-x).$$

Hence

$$A^3 = C^p = (AB)^2 = (BC)^2 = (CA)^2 = (ABC)^2 = 1;$$

that is, all the defining relations for  $G^{3,n,p}$  are satisfied, save that  $n$ , the period of  $B$ , remains to be determined. The matrix of  $B$ , namely

$$\begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix},$$

having the characteristic equation

$$X^2 + X - 1 = 0,$$

is similar to\*

$$\begin{pmatrix} -\tau & 0 \\ 0 & -\tau' \end{pmatrix}.$$

We are thus led to consider the substitution

$$\begin{pmatrix} \tau x \\ \tau' \end{pmatrix}$$

operating in the field of algebraic integers  $a+b\tau$ . We seek the smallest  $n$  for which

$$(\tau/\tau')^n \equiv 1 \pmod{p},$$

that is,

$$\tau^n \equiv \tau'^n \pmod{p},$$

$$5^{1/2}f_n \equiv 0 \pmod{p},$$

or (since  $f_n$  is a rational integer),

$$f_n \equiv 0 \pmod{p}.$$

Such a number  $n$  can always be found; in fact, by a known property of the Fibonacci numbers,  $n$  is a divisor of either  $p+1$  or  $p-1$  (except when  $p=5$ ). With this definition for  $n$ , the given substitutions  $A, B, C$  satisfy all the defining relations for  $G^{3,n,p}$ .

Now, it is well known† that  $A$  and  $C$  (or  $CA$  and  $C$ ) generate the linear fractional group  $\mathfrak{P}_1(p)$ . The remaining substitution  $B$ , having determinant  $-1$ , belongs to this group, or enlarges it to  $\tilde{\mathfrak{P}}_1(p)$ , according as  $-1$  is a quadratic residue or nonresidue, that is, according as  $p \equiv 1$  or  $3 \pmod{4}$ . In the latter case,  $n$  is necessarily even; so  $A$  and  $C$  satisfy the defining relations for  $(2, 3, p; n/2)$ . Hence we have the following theorem:

**THEOREM G.** *If  $p$  is a prime, congruent to 1 or 3 (mod 4), the group  $\mathfrak{P}_1(p)$  or  $\tilde{\mathfrak{P}}_1(p)$  (respectively) is a factor group of  $G^{3,n,p}$ , where  $n$  is the ordinal of the first Fibonacci number that is divisible by  $p$ . When  $p \equiv 3 \pmod{4}$ ,  $\mathfrak{P}_1(p)$  is a factor group of  $(2, 3, p; n/2)$ .*

Here are the first nine values of  $n$ :

\* For this method for calculating  $n$ , I am indebted to H. Hasse.

† Dickson [1], pp. 300-302.

$p$	3	5	7	11	13	17	19	23	29	...
$n$	4	5	8	10	7	9	18	24	14	...

Since  $A$  and  $C$  generate  $\mathfrak{P}_1(p)$ ,  $B$  must be expressible in terms of  $A$  and  $C$  whenever  $p \equiv 1 \pmod{4}$ . In fact, since

$$\begin{aligned}(a^2x) &= \left(-\frac{1}{x}\right)(x+a)\left(-\frac{1}{x}\right)\left(x+\frac{1}{a}\right)\left(-\frac{1}{x}\right)(x+a) \\ &= CA \cdot C^a \cdot CA \cdot C^{1/a} \cdot CA \cdot C^a = A^{-1}C^aAC^{1/a}A^{-1}C^{a-1},\end{aligned}$$

we have, when  $p \equiv 1 \pmod{4}$ ,

$$AB = (-x) = A^{-1}C^iAC^{-i}A^{-1}C^{i-1},$$

where  $i^2 \equiv -1 \pmod{p}$ . Thus

$$(3.92) \quad B = AC^iAC^{-i}A^{-1}C^{i-1}.$$

In Todd's definition\* for  $\mathfrak{P}_1(p)$ ,

$$S = C, \quad U = A, \quad R = A^{-1}C^aAC^{1/a}A^{-1}C^a.$$

When  $p=3, 5$ , or  $7$ ,  $\mathfrak{P}_1(p)$  or  $\tilde{\mathfrak{P}}_1(p)$  is the *whole* group  $G^{3,n,p}$  (see §3.2, §3.3). But, as we saw in §3.7,

$$(3.93) \quad \mathfrak{P}_1(13) \sim G^{3,7,13}/G_2.$$

It seems unlikely that  $G^{3,n,p}$  is finite in any higher case.

We observe that  $n$  is even when  $p=29$ . In this case, then, (3.92) (with  $i=12$ ) cannot be a consequence of the defining relations for  $G^{3,n,p}$ , since each of those relations involves  $B$  an even number of times. But I do not believe that this extra relation will suffice to reduce  $G^{3,14,29}$  to a finite group.

#### CHAPTER IV. GRAPHICAL REPRESENTATION

**4.1. Dyck's general group picture.** Dyck† represents the group

$$(4.11) \quad S_1^{b_1} = S_2^{b_2} = \cdots = S_r^{b_r} = S_1S_2 \cdots S_r = 1$$

by a "group picture" (*Dycksche Gruppenbild*) which consists of a network of  $r$ -gons, each of angles

$$\pi/b_1, \pi/b_2, \dots, \pi/b_r,$$

filling the sphere, the euclidean plane, or the hyperbolic plane, according as

\* Todd [2], p. 195. The period of  $R$  is, of course,  $(p-1)/2$ .

† Dyck [1], p. 26; Burnside [1], pp. 372-427; Threlfall [1], p. 26.

$$\frac{1}{b_1} + \frac{1}{b_2} + \cdots + \frac{1}{b_r}$$

is greater than, equal to, or less than,  $r-2$ . The polygons are supposed to be colored alternately white and black, and each side separates two polygons (of opposite colors) which are images of one another by reflection in that side. The group is transitive on the polygons of either color, and by adjoining one of the reflections we derive the *extended* group

$$(4.12) \quad \begin{cases} R_1^2 = R_2^2 = \cdots = R_r^2 = 1, \\ (R_r R_1)^{b_1} = (R_1 R_2)^{b_2} = \cdots = (R_{r-1} R_r)^{b_r} = 1, \end{cases}$$

for which the polygon is the fundamental region. The generators  $R_1, R_2, \dots, R_r$  are reflections in the sides of the fundamental region. The  $S$ 's are their products in consecutive pairs. The groups are finite when the representation is on a sphere, in which case  $r=2$  or  $3$  and (when  $r=3$ ) we have the relation  $1/b_1 + 1/b_2 + 1/b_3 > 1$ .

For many purposes, it is desirable to abandon the metric, and to regard the group picture as a network of polygons on a topological surface. We can then derive a group picture for any factor group, given by the insertion of extra relations in the abstract definition, by making the appropriate identification of various parts of the surface. (Such identification naturally reduces a simply connected surface to a multiply connected one.)

To obtain a group picture for the group derived from (4.12) by adding the relation

$$(4.121) \quad (R_1 R_2 \cdots R_r)^p = 1,$$

we should identify, in the group picture for (4.12) itself (that is, the "universal covering surface") all those polygons which represent operators of the subgroup generated by  $(R_1 R_2 \cdots R_r)^p$  and its conjugates. If  $rp$  is odd, this will involve the identification of a white region with a black one, so that the new surface will be unorientable. (Dyck would have represented the same group on the orientable twofold covering surface.)

**4.2. The group picture for  $G^{m,n,p}$ .** In the case when  $r=3$  and  $b_1, b_2, b_3$  are  $2, m, n$ , we have a factor group of  $[m, n]$ , namely

$$R_1^2 = R_2^2 = R_3^2 = (R_1 R_2)^m = (R_2 R_3)^n = (R_3 R_1)^2 = (R_1 R_2 R_3)^p = 1.$$

We saw, in §3.2, that this is precisely  $G^{m,n,p}$ , the ordinary definition being given by the substitution

$$(4.21) \quad \begin{array}{lll} R_1 = BC, & R_2 = BCA, & R_3 = CA, \\ A = R_1 R_2, & B = R_2 R_3, & C = R_3 R_2 R_1. \end{array}$$

When  $p$  is even, our group picture for  $G^{m,n,p}$  is Dyck's for the subgroup  $(2, m, n; p/2)$ , which is generated by  $A, B$  (and  $AB$ ). When  $p$  is odd,  $A$  and  $B$  generate the whole group  $G^{m,n,p}$ , and the distinction between white and black evaporates.

Klein\* shows a stereographic projection of the group picture for  $[3, 5]$ . The sides of the triangles appear as arcs of circles, meeting at the proper

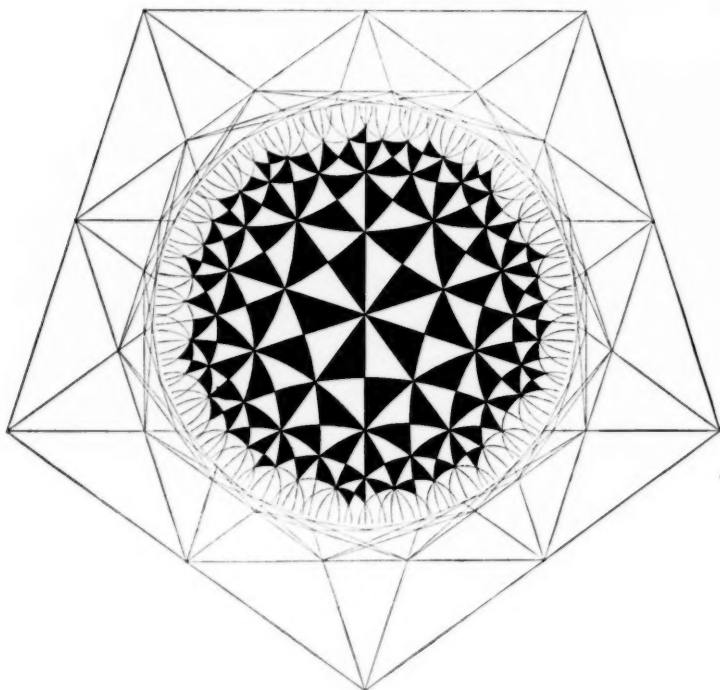


FIG. 6

The group picture for  $[4, 5]$

angles, and the reflections are represented by inversions in these circles. Elsewhere† he and Fricke show the analogous conformal representation of the (hyperbolic) group picture for  $[3, 7]$ . For  $[4, 5]$  and  $[4, 6]$ , see Figs. 6 and 7, where I have shown also the construction lines for the centers of the circles.

\* Klein [1], p. 260.

† Klein-Fricke [1], p. 109, Fig. 33.

(Since the circles belong in various ways to coaxial systems, their centers lie conveniently on straight lines. In the case of  $[4, 6]$ , all the centers are derivable from the outermost hexagon in a remarkably simple manner.)

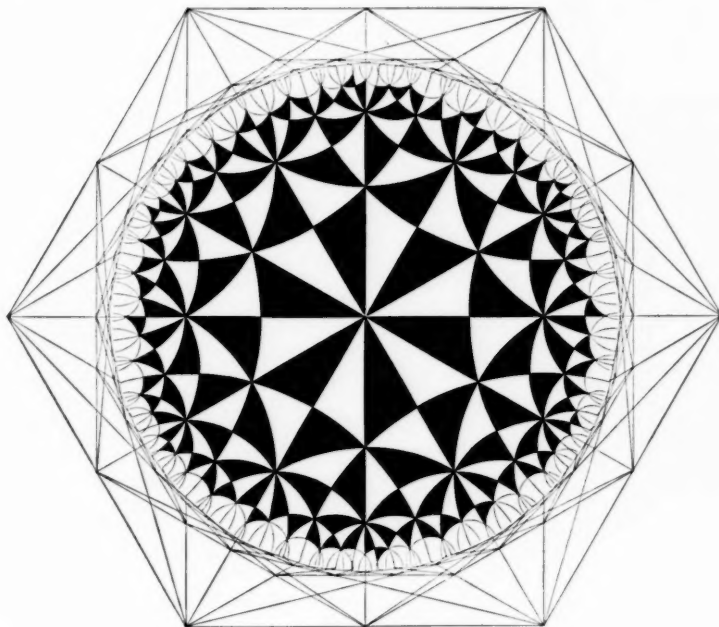


FIG. 7  
The group picture for  $[4, 6]$

4.3. **The regular map  $\{m, n\}_p$ .** Clearly, the vertices fall into three categories, say  $P_2$ ,  $P_m$ ,  $P_n$ , according as the number of triangles of either color that surround the vertex is 2,  $m$ , or  $n$ . The  $2m$  triangles that surround one point  $P_m$  form together a regular  $m$ -gon,  $\{m\}$ , and the totality of such  $m$ -gons constitutes a *regular map*\*  $\{m, n\}$ , whose vertices (each surrounded by  $n$   $\{m\}$ 's) are the points  $P_n$ . Similarly, the points  $P_m$  are the vertices of the reciprocal map  $\{n, m\}$ . The points  $P_2$  are the mid-points of the edges of either map. In the case when  $m=n$ , the points  $P_m$  and  $P_n$  are surrounded alike; taken all together, they are the vertices of a regular map of quadrangles,  $\{4, n\}$ .

\* Brahana [1]. This is the *regelmässige Zellsystem*  $\{n, m\}$  of Threlfall [1], p. 32. Following Schläfli, van Oss, and Sommerville, I write the  $m$  (or  $a_2$ ) before the  $n$  (or  $a_0$ ).

Let  $P, Q, R, S, T, \dots$  be a sequence of vertices of  $\{m, n\}$ , so chosen that  $P, Q, R$  are consecutive vertices of one  $m$ -gon,  $Q, R, S$  of another,  $R, S, T$  of another, and so on. The edges  $PQ, QR, RS, \dots$  form a kind of zig-zag, which we call a *Petrie polygon*. (See Fig. 8, for the case of  $\{5, 4\}$ .) We easily see that the operator  $R_1 R_2 R_3$  cyclically permutes the vertices (and sides) of a Petrie polygon. Consequently, the insertion of the relation  $(R_1 R_2 R_3)^p = 1$  corresponds to the identification of all pairs of vertices which are separated by  $p$  consecutive sides of a Petrie polygon. The reduced map, so derived, will be denoted by  $\{m, n\}_p$ . The map  $\{5, 10\}_3$  is one of Brahana's and Coble's unorientable dodecahedra.\*

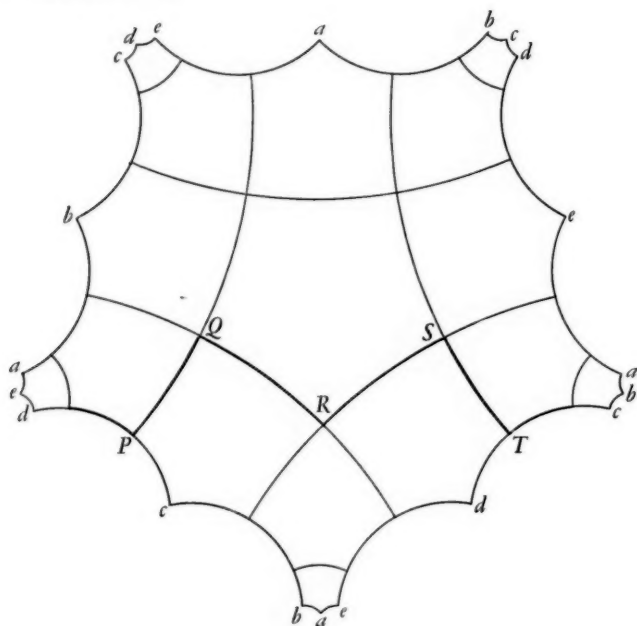


FIG. 8  
The regular map  $\{5, 4\}_3$

**4.4. The semi-regular map  $\{m/n\}_p$ .** In the metrical representation of  $\{m, n\}$ , the mid-points of the sides of the Petrie polygon lie on a straight line (of the euclidean or non-euclidean plane) or a circle (of the stereographic projection). This suggests the desirability of replacing the regular map by a *semi-regular map*, whose vertices are all the points  $P_2$ .

\* Brahana and Coble [1], p. 15, Fig. VIII.

This semi-regular map, which we denote by  $\{m/n\}$  (or  $\{n/m\}$ ), has faces of two types,  $\{m\}$  and  $\{n\}$ , whose centers are the points  $P_m$  and  $P_n$ , respectively. (When  $m \neq n$ , these faces are actually different; when  $m = n$ , they can still be distinguished, like the black and white squares of a checkerboard. See Fig. 11.) Every  $\{m\}$  is surrounded by  $m$   $\{n\}$ 's, every  $\{n\}$  by  $n$   $\{m\}$ 's, and every vertex by two of each, alternating. Moreover, when an edge is produced in both directions, the line obtained (corresponding to the Petrie polygon of  $\{m, n\}$ ) is entirely covered with edges.

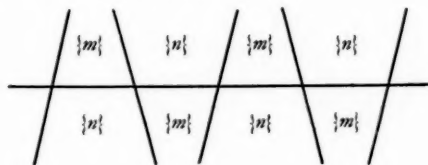


FIG. 9  
Collinear edges of  $\{m/n\}$

The operator  $R_1 R_2 R_3$  permutes the edges by one step along such a line, and the relation  $(R_1 R_2 R_3)^p = 1$  corresponds to the identification of all pairs of vertices which differ by  $p$  such steps. When this identification has been effected, we denote the semi-regular map by\*  $\{m/n\}_p$ , reserving the symbol  $\{m/n\}$  itself for the universal covering map (which is  $\{m/n\}_\infty$  unless  $1/m + 1/n > 1/2$ ).

As we have already remarked, the map is unorientable when  $p$  is odd. In fact, since the collinear edges have alternate  $\{m\}$ 's and  $\{n\}$ 's on the left, and vice versa on the right, the surface in the neighborhood of the line forms a "Möbius band" (see Fig. 9).

When  $m = n$ , the map  $\{m/n\}_p$  is regular instead of semi-regular. In particular,  $\{m/m\}$  is  $\{m, 4\}$ , and  $\{m/m\}_{2q}$  is the "skew polyhedron"†  $\{m, 4 | 2q\}$ , whose rotation group is  $(m, 4 | 2, 2q)$  (see Fig. 18). We must *not* infer, however, that

$$G^{m, m, 2q} \sim (m, 4 | 2, 2q),$$

although this happens to be true when  $m = 3$  or  $5$  and  $q = 2$  (see (3.31)). For the map represents the two groups in quite different ways:  $(m, 4 | 2, 2q)$  contains no reflections, but  $G^{m, m, 2q}$  (containing reflections) will not transform an  $m$ -gon into any adjacent  $m$ -gon. All that we may infer is that the two groups

\* Since  $\{m/n\}_p$  has the same meaning as  $\{n/m\}_p$ , perhaps a better symbol would have been  $\{n/m\}_p$ . The symbol  $\{n/m\}$ , here written  $\{m/n\}$ , was defined in Coxeter [1], p. 127.

† Coxeter [7], p. 50.

have a common subgroup of index two, as we already know from (2.51) and §3.1, the subgroup being  $(2, m, m; q)$ .

Since  $G^{m,n,p}$  involves  $m, n, p$  symmetrically, the three maps  $\{m/n\}_p$ ,  $\{n/p\}_m$ ,  $\{p/m\}_n$  all represent the same group. (If  $m \leq n \leq p$ , the first of the three is usually the most convenient, as having the lowest connectivity.) The group  $G^{m,n,p}$  is also represented by any of the six corresponding regular maps

$$\{m, n\}_p, \quad \{n, m\}_p, \quad \{n, p\}_m, \quad \{p, n\}_m, \quad \{p, m\}_n, \quad \{m, p\}_n.$$

The map  $\{m, 4\}_m$  has four  $\{m\}$ 's at each vertex, just like  $\{m/m\}_4$ ; it represents the same group  $G^{4,m,m}$ , but has half as many  $\{m\}$ 's altogether. In fact,  $\{m/m\}_4$  is its twofold covering surface. (See Figs. 8 and 18. The outermost edges are supposed to be identified in accordance with the lettering.)

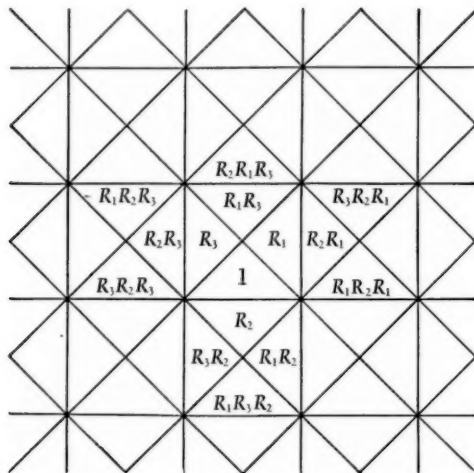


FIG. 10  
 $G^{4,4,4}$  as a factor group of  $[4, 4]$

**4.5. The manner in which  $\{m/n\}_p$  represents  $G^{m,n,p}$ .** Since each edge of  $\{m/n\}_p$  crosses the common hypotenuse of two of Dyck's triangles (one white and one black), it is the half-edges that represent the operators of  $G^{m,n,p}$ . When  $p$  is even, the edges themselves represent the operators of the subgroup  $(2, m, n; p/2)$ .<sup>\*</sup> Hence, if  $g$  denotes the order of  $G^{m,n,p}$ , the semi-

<sup>\*</sup> When both  $n$  and  $p$  are even, the edges of  $\{m/n\}_p$  can be indexed so as to form a Cayley color-group (Burnside [1], pp. 423–427) or *Dehnsche Gruppenbild* (Threlfall [1], pp. 22–27) for  $(m, m|n/2, p/2)$ , thus providing a geometrical interpretation for (2.52).

regular map  $\{m/n\}_p$  has  $g/2$  edges,  $g/4$  vertices,  $g/2m$   $\{m\}$ 's, and  $g/2n$   $\{n\}$ 's. Its Euler-Poincaré characteristic\*  $(-V+E-F)$  is therefore

$$\frac{g}{2} \left( \frac{1}{2} - \frac{1}{m} - \frac{1}{n} \right).$$

The generators  $A, B, C$  are to be interpreted as follows.  $A$  is an  $m$ -gonal rotation about the center of a face  $\{m\}$ ;  $B$  is an  $n$ -gonal rotation (in the same sense) about the center of an adjacent face  $\{n\}$ ; and  $C$ , being the inverse of

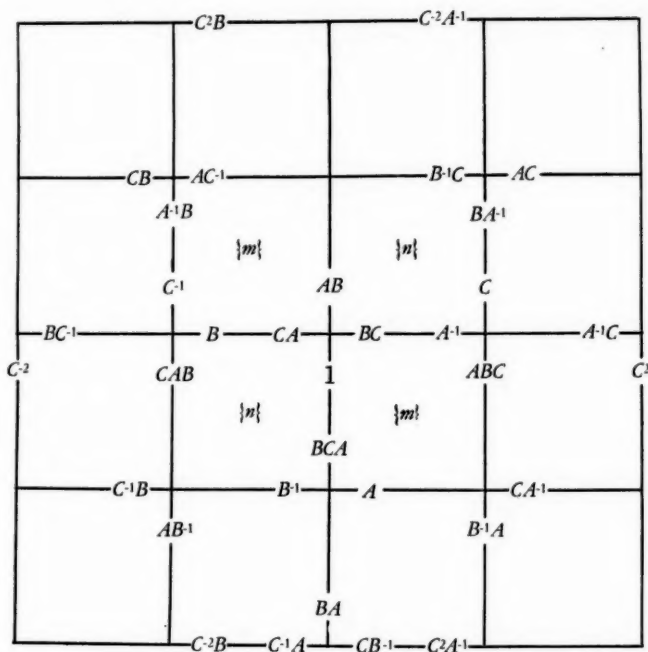


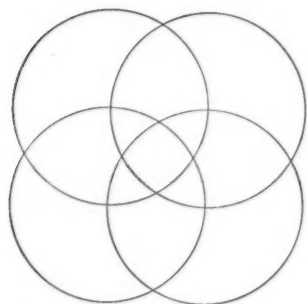
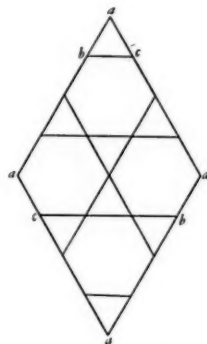
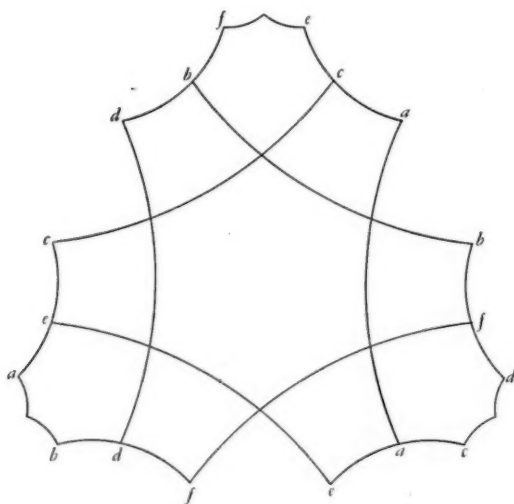
FIG. 11  
 $\{4/4\}_4$  as a portion of  $\{4/4\}$

$R_1 R_2 R_3$ , is a "glide" along a chain of collinear edges. To be precise,  $C$  transforms the edge one half of which represents  $B$  into that whose "other" half represents  $A^{-1}$ .

4.6. **Special cases.** Figs. 10 and 11 show the two kinds of group picture for  $G^{4,4,4}$  (of order 64); namely, Dyck's picture for  $(2, 4, 4; 2)$ , and the map

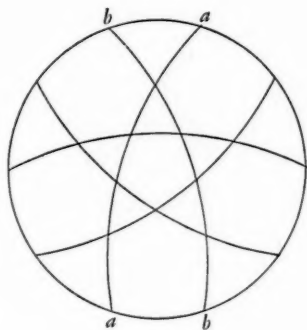
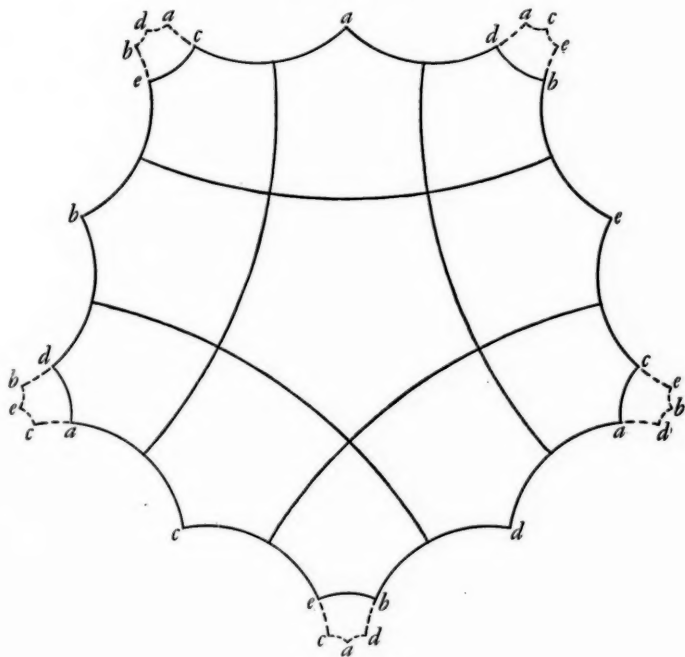
\* An orientable surface of characteristic  $c$  has genus  $c/2+1$ .

$\{4/4\}_4$ . Some of the operators have been explicitly inserted; the rest are easily deduced. The generalization to  $G^{4,4,p}$  for any even  $p$  is obvious, as is the collapse of  $G^{4,4,p}$  for any odd  $p$ .

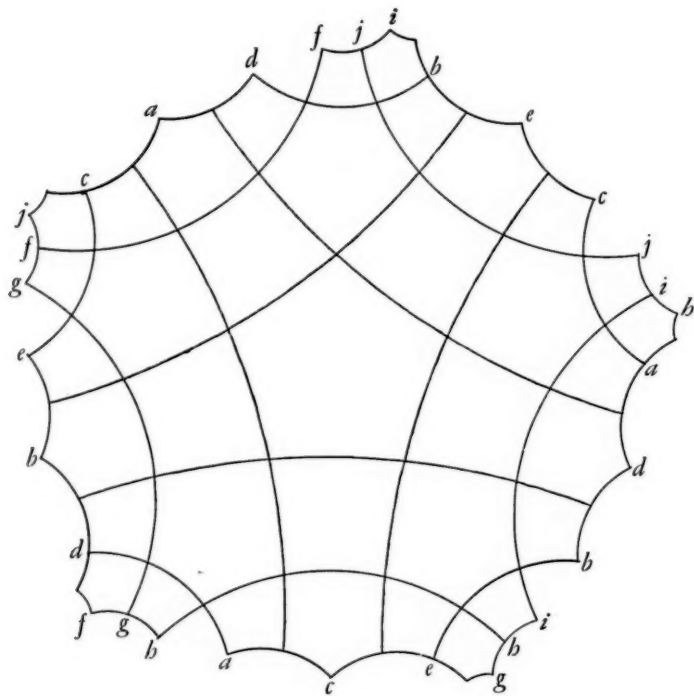
FIG. 12.  $\{3/4\}_4$ FIG. 13.  $\{3/6\}_4$ FIG. 14.  $\{4/6\}_3$ 

The extended octahedral group  $G^{3,4,6}$  provides an interesting example, as its three pictures (Figs. 12, 13, 14) are, respectively, spherical, euclidean, and hyperbolic. The map  $\{3/6\}_4$  generalizes at once to  $\{3/6\}_p$  for any even  $p$ . This can always be drawn as a rhombus (of angle  $\pi/3$ ), with opposite sides

identified, like the period parallelogram of an elliptic function.\* By trying to make an analogous picture when  $p$  is odd, we verify the collapse of  $G^{3,p}$  ( $p$  odd).

FIG. 15.  $\{3/5\}_8$ FIG. 16.  $\{5/5\}_8$ 

\* Edington [1], p. 206.

FIG. 17.  $\{4/5\}_8$ 

Figs. 15 and 16 represent the icosahedral group  $G^{3,5,5}$ . In the former, all pairs of opposite points of the outermost circle are to be identified. If all six circles were completed, we should have  $\{3/5\}_{10}$  or  $\{3/5\}$ . The map  $\{3/5\}_8$  may be regarded as a partition of the elliptic plane into ten triangles and six pentagons;  $\{3/5\}$  is its twofold covering surface. Similarly,  $\{3/4\}$  is the twofold covering surface of  $\{3/4\}_3$ , which may be regarded as a partition of the elliptic plane into four triangles and three squares.

In Fig. 16, the smallest pentagon has been drawn five times over (in broken lines), to preserve the pentagonal symmetry of the figure. (The re-

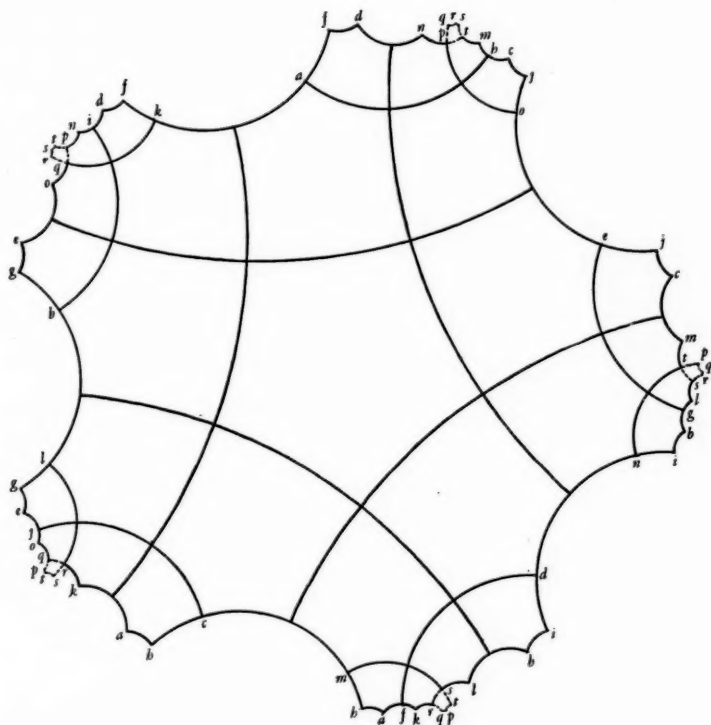


FIG. 18  
 $\{5/5\}_4$  or  $\{5, 4 \mid 4\}$

semblance to Fig. 8 is therefore spurious.)  $\{5/5\}_3$  is another of Brahana's and Coble's unorientable dodecahedra.\* Its twofold covering surface is  $\{5, 4\}_6$ .

Figs. 17 and 18 represent the group  $G^{4,5,5}$ , of order 160. These are unorientable and orientable, respectively. The latter is the twofold covering sur-

\* Brahana and Coble [1], p. 3 (Fig. II). Similarly,  $\{5, 6\}_4$  is the twofold covering surface of Fig. III (p. 9).

face of the unorientable map shown in Fig. 8. It represents also the subgroup  $(2, 5, 5; 2)$ , of order 80, which is the abelian group of order sixteen and type  $(1, 1, 1, 1)$  augmented by an operator of period five.

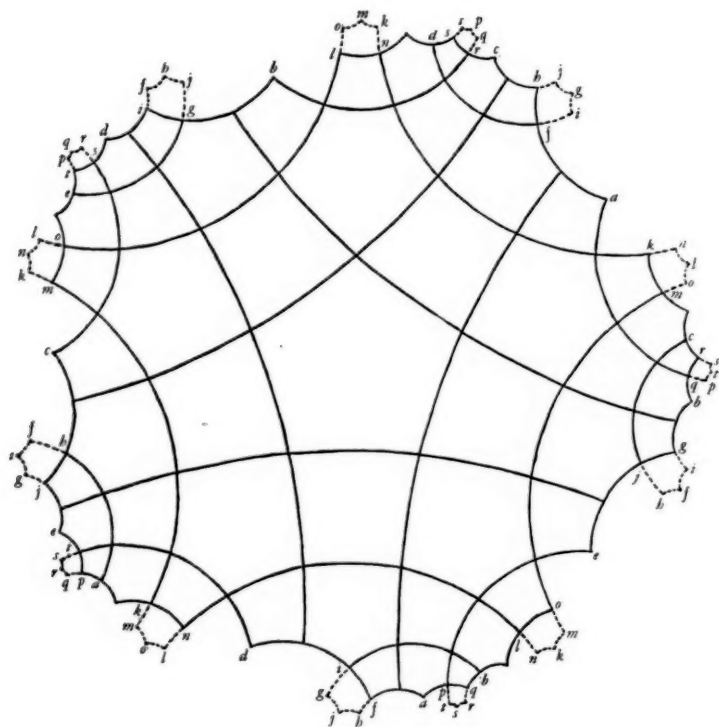


FIG. 19.  $\{4/5\}_6$

Figs. 19 and 20 show two of the semi-regular maps for  $G^{4,5,6}$ . The former, being orientable, represents also the subgroup  $(2, 4, 5; 3)$ , which is the sym-

metric group of degree five. (In Fig. 20, the small hexagon  $mnpqr$  has not been repeated, although it upsets the symmetry of the figure.)

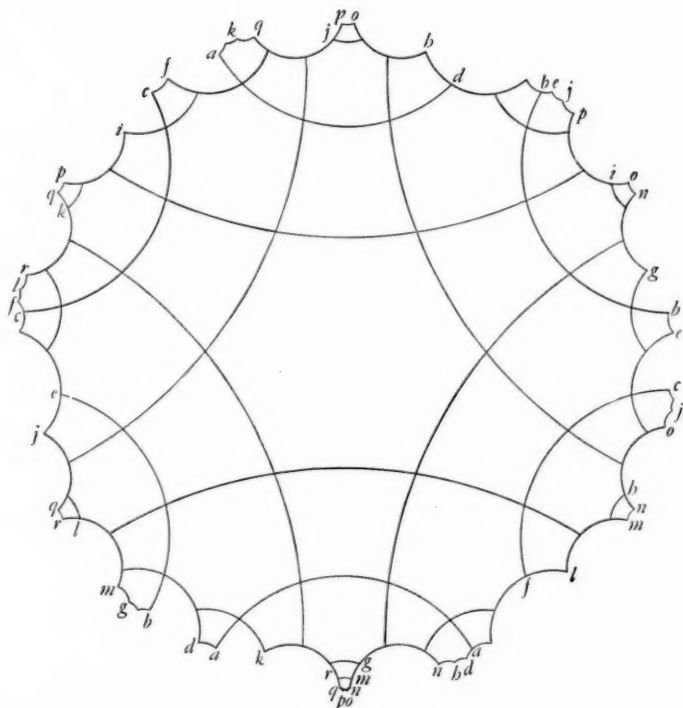


FIG. 20.  $\{4/6\}_s$

Fig. 21 represents  $G^{3,7,8}$ , the 168 edges representing the operators of the subgroup  $(2, 3, 7; 4)$ , which is  $\mathfrak{P}_1(7)$ .

Finally, Fig. 22 represents  $G^{3,7,9}$ , which is  $\mathfrak{P}_1(8)$ , the simple group of order 504. The peculiar elegance of this figure is partly due to the fact that

the number of heptagons, being the number of Sylow subgroups of order seven, is one more than a multiple of seven, so that the heptagonal symmetry

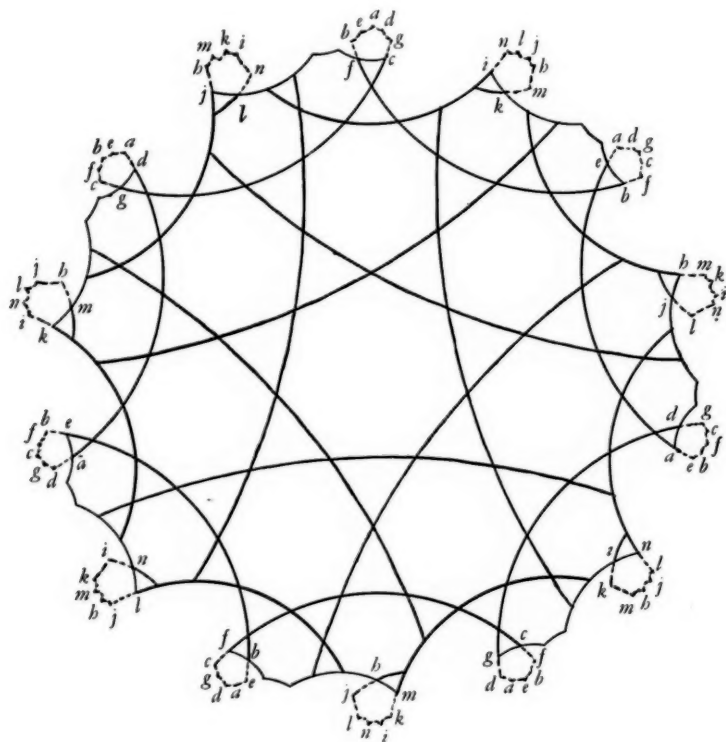


FIG. 21.  $\{3/7\}_8$

can be preserved without repeating any of the faces. (In Fig. 21, there are *three* heptagons for every such subgroup of  $G^{3,7,8}$ .) Fig. 22 incidentally provides a geometrical verification for the collapse\* of  $G^{3,7,7}$ . The numbers

\* Cf. Sinkov [5].

1, 2, 3, 4, 5 have been inserted in four of the heptagons, each consecutive pair of which differ by seven steps along a line of collinear edges. To make

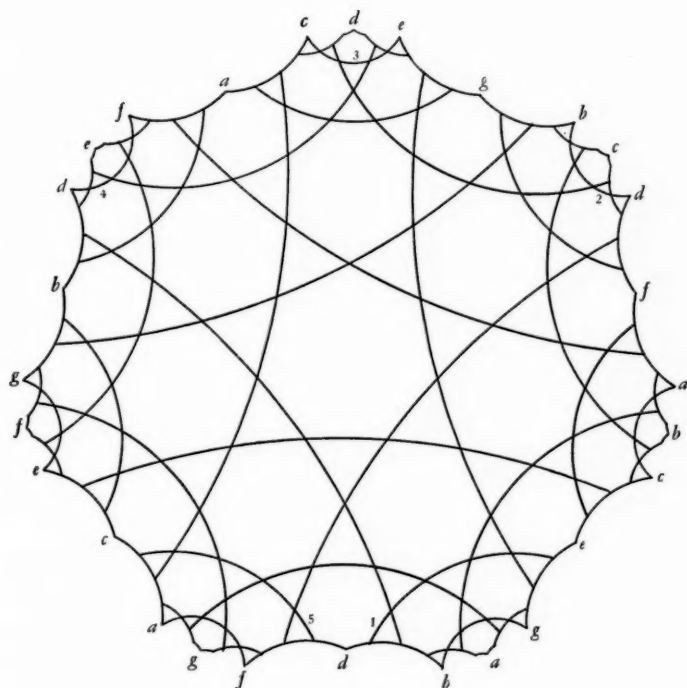


FIG. 22.  $\{3/7\}_9$

$\{3/7\}_7$ , these heptagons would have to be identified. But the numbers have been put against particular sides of the heptagons, namely, sides which would have to be identified. We are thus led to the identification of two sides (1 and 5) of one heptagon, which is absurd.

4.7. **The polyhedral groups.** When  $1/m + 1/n > 1/2$ , the vertices of the maps

$$\{m, n\}, \quad \{n, m\}, \quad \{m/n\},$$

regarded as points of a sphere, can be joined by straight lines and planes (in ordinary space) to form regular and semi-regular polyhedra. From this point of view,\*  $\{3, 3\}$  is the tetrahedron,  $\{3/3\}$  or  $\{3, 4\}$  is the octahedron,  $\{4, 3\}$  is the cube,  $\{3/4\}$  is the cuboctahedron,†  $\{3, 5\}$  is the icosahedron,  $\{5, 3\}$  is the dodecahedron, and  $\{3/5\}$  is the icosidodecahedron.‡

When  $\{m/n\}$  is regarded as a polyhedron, the "collinear edges" form an equatorial§  $p$ -gon,  $C$  being a rotary reflection whose period  $p$  we proceed to calculate. Consider the four vertices which are joined by edges to any one vertex. These form a rectangle, whose sides are  $2 \cos \pi/m$  and  $2 \cos \pi/n$ , while its diagonal is  $2 \cos \pi/p$ . Hence

$$\cos^2 \pi/p = \cos^2 \pi/m + \cos^2 \pi/n, \quad 1/m + 1/n \geq 1/2,$$

in agreement with (3.23).

This formula fails when it leads to an odd value for  $p$ , since then  $C^p$ , instead of being identity, is the reflection in the plane of the  $p$ -gon. This happens only when  $m=2$  and  $n$  is odd (or *vice versa*). In order to cover this case, we may say that the period of  $C$  is twice the numerator of the rational number  $q$ , defined by

$$(4.71) \quad \cos \pi/q = 1 + \cos 2\pi/m + \cos 2\pi/n.$$

4.8. **Polyhedra of higher connectivity.** From the ordinary polyhedron  $\{m/n\}$  or  $\{m/n\}_p$ , we may derive two *star polyhedra*,  $\{n/p\}_m$  and  $\{m/p\}_n$ , by regarding the equatorial  $p$ -gons as faces, and discarding the  $m$ -gons or  $n$ -gons, respectively. Thus we have the *tetratrihedron*||  $\{3/4\}_3$ , the *hexatetrahedron*¶  $\{4/6\}_3$ , the *octatetrahedron*¶  $\{3/6\}_4$ , the *dodecahexahedron*\*\*  $\{5/10\}_3$ , and the *icosihexahedron*\*\*  $\{3/10\}_6$ .

The cuboctahedron,  $\{3/4\}$  or  $\{3/4\}_6$ , represents the extended octahedral group  $G^{3,4,6}$ , which has two distinct subgroups of index two:  $(2, 3, 4; 3)$  and  $(2, 3, 6; 2)$ . The former appears geometrically as the octahedral group, generated by a trigonal rotation  $A$  and a tetragonal rotation  $B$ , while the latter

\* Cf. Coxeter [1], p. 129.

† Badoureau [1], p. 67, Fig. 30; p. 73, Fig. 39.

‡ Ibid., p. 73, Fig. 45; p. 129, Figs. 113, 114.

§ That is, inscribed in a great circle.

|| The *semi-octaèdre* of Badoureau [1], p. 104, Fig. 70.

¶ Ibid., p. 119, Figs. 96, 97.

\*\* Ibid., p. 131, Figs. 115, 116.

appears as the pyritohedral group, generated by the same  $A$  and a rotary reflection  $C$ . When the same group is represented on the octatetrahedron  $\{3/6\}_4$  (whose faces are the eight triangles of the cuboctahedron, and its four equatorial hexagons), the roles of the two subgroups are interchanged: the generators,  $A$  and  $C$ , of the pyritohedral subgroup  $(2, 3, 6; 2)$ , though still a rotation and a rotary reflection in space, are both *intrinsic* rotations\* of the surface of the polyhedron. On the other hand, the third polyhedron,  $\{4/6\}_3$  (whose faces are the six squares of  $\{3/4\}_6$  and the four hexagons of  $\{3/6\}_4$ ) is unorientable, and its intrinsic rotations,  $B$  and  $C$ , generate the whole (extended octahedral) group. Similarly, the intrinsic rotations of  $\{3/4\}_6$  generate the extended tetrahedral group  $G^{3,3,4}$  (which is simply isomorphic with the octahedral group), while those of  $\{5/10\}_3$ , or of  $\{3/10\}_6$ , generate the extended icosahedral group  $G^{3,5,10}$ .

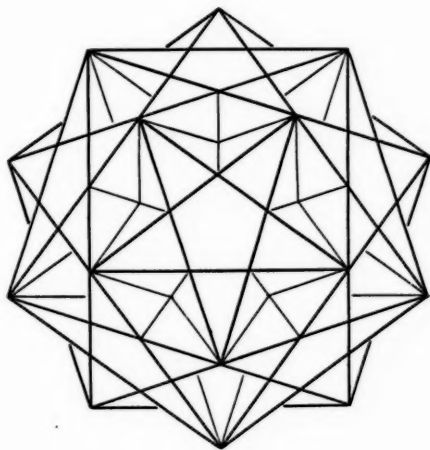


FIG. 23

The ditrigonal dodecadodecahedron, homeomorphic to  $\{5, 6\}_4$

The cuboctahedron may be said to be *semi-reciprocal* to the cube and the octahedron. Semi-reciprocal to the Kepler-Poinsot polyhedra  $\{5, \frac{5}{2}\}$  and  $\{\frac{5}{2}, 5\}$  there is the dodecadodecahedron†  $\{5/\frac{5}{2}\}$ , which has the same vertices as the icosidodecahedron, but its faces consist of twelve pentagons and twelve pentagrams (or star pentagons). By ignoring the distinction between

\* Cf. Coxeter [7], p. 47.

† Badoureau [1], p. 133, Fig. 117.

these two kinds of face (which, topologically speaking, are both pentagons), we derive the map  $\{5, 4\}_6$ , whose intrinsic rotation group,  $(2, 5, 4; 3)$ , is  $G_{51}$ . In fact, the pentagons and pentagrams of  $\{5 / \frac{5}{2}\}$  are interchanged by any outer automorphism of the icosahedral group (which is the rotation group of the solid  $\{5 / \frac{5}{2}\}$ ).

All this is closely analogous to the manner in which two reciprocal tetrahedra  $\{3, 3\}$  lead to the octahedron (or "tetratetrahedron")  $\{3/3\}$ , which is the same polyhedron as  $\{3, 4\}_6$ . For, the octahedral group  $(2, 3, 4; 3)$  is the group of isomorphisms of the tetrahedral group.

Besides the dodecadodecahedron  $\{5 / \frac{5}{2}\}$ , there is another polyhedron (Fig. 23) having the same number of the same kinds of face (namely, pentagons and pentagrams), but three of each at a vertex, instead of only two. We may call this the *ditrigonal dodecadodecahedron*.\* It has the same vertices as the ordinary dodecahedron. By ignoring the distinction between pentagons and pentagrams, we derive the map  $\{5, 6\}_4$ , whose intrinsic rotation group,  $(2, 5, 6; 2)$ , is again  $G_{51}$ .

**4.9. The more general group  $((b_1, b_2, \dots, b_r; p))$ .** By analogy with (3.15), we use the symbol  $((b_1, b_2, \dots, b_r; p))$  to denote the group defined by (4.12) and (4.121). The substitution (4.21) enables us to define  $((l, m, n; p))$  in the alternative form

$$(4.91) \quad A^m = B^n = C^p = (AB)^l = (BC)^2 = (CA)^2 = (ABC)^2 = 1,$$

which shows that

$$(4.92) \quad ((2, m, n; p)) \sim G^{m,n,p}.$$

Three methods present themselves for a tentative investigation of the group  $((b_1, b_2, \dots, b_r; p))$ . The first applies to the case when  $r=3$  and  $p$  is even. We saw, in §3.1, that  $((l, m, n; 2q))$  contains  $((l, m, n; q))$  as a subgroup of index two. By writing

$$R = R_3 R_1, \quad S = R_1 R_2, \quad T = R_2 R_3,$$

so that

$$(R_1 R_2 R_3)^2 = SRT,$$

we obtain the subgroup in the symmetrical form (2.111), that is, as a factor group of (4.11) (with  $r=3$ ). We may suppose  $l, m, n$  to be greater than 2, since  $G^{m,n,p}$  has already been thoroughly investigated.

From  $(bcd, cad, abd; 1)$  ( $a, b, c$  co-prime in pairs) we derive the

\* Badoureaux [1], p. 108, Fig. 76, incomplete.

group  $((bcd, cad, abd; 2))$ , of order  $2abcd^2$ ,  
 from  $(3, 3, 3; k)$ ,  $((3, 3, 3; 2k))$ , of order  $18k^2$ ,  
 from  $(3, 3, 12; 2)$ ,  $((3, 3, 12; 4))$ , of order 576,  
 from  $(3, 3, 6; 2)$ ,  $((3, 3, 6; 4))$ , of order 288,  
 from  $(3, 3, 4; 3)$ ,  $((3, 3, 4; 6))$ , of order 1008,  
 from  $(3, 4, 4; 2)$ ,  $((3, 4, 4; 4))$ , of order 480,  
 from  $(3, 4, 5; 2)$ ,  $((3, 4, 5; 4))$ , of order 720.\*

We see also that  $((3, 3, n; 4))$  collapses unless  $n=2, 3, 6$ , or 12, and that  $((4, 4, 4; 4))$  and  $((3, 3, 4; 8))$  are infinite.

The second method applies to the case when

$$b_1 = b_2 = \dots = b_r = q \text{ (say).}$$

By adjoining an operator  $S$ , of period  $r$ , which cyclically permutes the generators  $R_1, R_2, \dots, R_r$ , we derive the larger group

$$S^r = R_1^2 = (SR_1)^{r^2} = (S^{-1}R_1SR_1)^q = 1.$$

Thus†  $((q^r; p))$  is an invariant subgroup of index  $r$  in  $(2, r, rp; q)$ .

From  $(2, 2, 2p; p) \sim [2p]$ , we derive  $((p, p; p)) \sim [p]$ , of order  $2p$ ,

from  $(2, 3, 3; 2)$ , the four-group  $((2, 2, 2; 1))$ ,  
 from  $(2, 3, 6k; 3)$ ,  $((3, 3, 3; 2k))$  again,  
 from  $(2, 3, 6; q)$ ,  $((q, q, q; 2))$ , of order  $2q^2$ ,  
 from  $(2, 4, 4; q)$ ,  $((q, q, q, q; 1))$ , of order  $2q^2$ ,  
 from  $(2, 4, 4p; 2)$ ,  $((2, 2, 2, 2; p))$ , of order  $8p^2$ ,  
 from  $(2, 5, 5; 2)$ ,  $((2, 2, 2, 2, 2; 1))$ , of order 16,

and from  $(2, 3, 12; 4)$ , the infinite group  $((4, 4, 4; 4))$ , again. We see also that the following are cases of collapse:‡

$$((q, q; p)) \quad (p \neq q), \quad ((q, q, q; 1)) \quad (q \neq 2),$$

$$((2, 2, 2; p)) \quad (p > 2), \quad ((3, 3, 3; p)) \quad (p \text{ odd}), \text{ and } ((4, 4, 4; 3)).$$

Theorem C tells us that  $((q, q, q, q; p))$  is infinite whenever  $p > 1$  or  $q > 2$ , and that  $((q^r; 1))$  is infinite when  $r > 4$ , with the single exception of  $((2, 2, 2, 2, 2; 1))$ . When  $r \leq 5$ ,  $((2^r; 1))$  is the abelian group of order  $2^{r-1}$  and type  $(1, 1, \dots, 1)$ . This is obvious when  $r < 5$ ; however, when  $r = 5$  it is an in-

\* The method of §3.3, applied to the permutations (2.92), shows immediately that  $((3, 4, 5; 4)) \sim G_{81/2} \times G_2$ .

† Within the double parentheses, we use  $q^r$  to stand for  $q, q, \dots, q$ .

‡ Enumerations, carried out independently by Sinkov and me (August–September, 1938), indicate the collapse of  $((5, 5, 5; 3))$ , and so also of  $(2, 3, 9; 5)$  and  $G^{3,9,10}$  (see Fig. 3). This is particularly interesting, as it shows that (3.81) is not a necessary condition for the collapse of  $G^{m,n,p}$ .

interesting consequence of our knowledge of the order of the group  $(2, 5, 5; 2)$ .

The  $b$ 's being all equal (to  $q$ ), the group picture described in §4.1 is now a regular map of  $r$ -gons,  $2q$  at each vertex. Since the relation (4.121) identifies pairs of points which differ by  $rp$  steps along a Petrie polygon, this map is precisely  $\{r, 2q\}_{rp}$ . For instance, the sixteen pentagons of Fig. 8 represent the operators of the abelian group of order sixteen and type  $(1, 1, 1, 1)$ .

The third method applies to the case  $r=3$ , without restriction on the parity of  $p$ ; but it is practical only when  $p=3$  or  $4$ .<sup>\*</sup> By comparing (4.91) with

$$T_1^{k_1} = T_2^{k_2} = T_3^{k_3} = (T_1 T_2)^2 = (T_2 T_3)^2 = (T_1 T_2 T_3)^2 = 1,$$

which are Todd's† relations, we see that, for a suitable value of  $l$ ,

$$((l, m, n; p)) \sim [n, p, m]',$$

where  $[n, p, m]'$  denotes the rotation group of the regular polytope  $\{n, p, m\}$  in four dimensions. Here  $l$  is the period of  $T_1 T_3$ , and  $T_1 T_3$  is a displacement‡ which cyclically permutes the sides (and vertices) of a Petrie polygon. In other words, the Petrie polygon of  $\{n, p, m\}$  has  $l$  sides.

By considering the particular polytopes in turn, we find§

$$\begin{aligned} ((3, 3, 5; 3)) &\sim [3, 3, 3]' \sim G_{51/2}, \\ ((3, 4, 8; 3)) &\sim [3, 3, 4]', \text{ of order 192,} \\ ((3, 3, 12; 4)) &\sim [3, 4, 3]', \text{ of order 576,} \\ ((3, 5, 30; 3)) &\sim [3, 3, 5]', \text{ of order 7200.} \end{aligned}$$

Moreover, since  $(T_1 T_3)^{l/2}$  ( $l$  even) is the central inversion,|| we have the central quotient groups¶

$$\begin{aligned} ((3, 4, 4; 3)) &\sim [3, 3, 4]'/G_2, \text{ of order 96,} \\ ((3, 3, 6; 4)) &\sim [3, 4, 3]'/G_2, \text{ of order 288,} \\ ((3, 5, 15; 3)) &\sim [3, 3, 5]'/G_2, \text{ of order 3600,} \end{aligned}$$

which can be regarded as rotation groups in elliptic space.¶¶

<sup>\*</sup> Added in proof: By direct enumeration of cosets, J. M. Kingston found that  $((3, 3, 4; 5)) \sim ((3, 3, 5; 5)) \sim G_{61/2}$ . As permutations, the generators are  $(2\ 3)(5\ 6)$ ,  $(1\ 2)(3\ 4)$ ,  $(1\ 2)(4\ 5)$  in the former, and  $(2\ 3)(4\ 5)$ ,  $(1\ 2)(3\ 4)$ ,  $(1\ 2)(4\ 6)$  in the latter.

† Todd [1], p. 217.

‡  $T_1 T_3 = R_1 R_2 R_3 R_4$ ; *ibid.*, p. 224. See also Coxeter [2], p. 605, where the period of  $R_1 R_2 R_3 R_4$  is called  $h$ .

§ For the sake of uniformity, I have rearranged  $l, m, n$  into ascending order. For example, the first group would originally appear as  $((5, 3, 3; 3))$ . Although the meaning of the symbol  $((l, m, n; p))$  is unchanged for all permutations of  $l, m, n$ , the meaning of  $[n, p, m]'$  is unchanged only for transposition of  $m$  and  $n$ . Moreover, the two polytopes  $\{n, p, m\}$ ,  $\{m, p, n\}$ , which have the same group, are not identical (unless  $m=n$ ) but *reciprocal*.

|| Coxeter [2], p. 606.

¶ Nos. XXVII, XXVIII, XXX of Goursat [1], pp. 67, 68.

We may also infer that  $((3, 3, l; 3))$  collapses if  $l \neq 5$ , and that  $((3, 4, l; 3))$  and  $((3, 5, l; 3))$  collapse whenever  $l$  is not a divisor of 8, or 30, respectively.\*

Each finite polytope  $\{n, p, m\}$  determines a partition of the hypersphere into polyhedral cells, which form a three-dimensional "regular map." More generally, we use the same symbol  $\{n, p, m\}$  to denote any regular map which fills a simply connected three-space (spherical, euclidean, or hyperbolic, according as  $\sin \pi/m \sin \pi/n$  is greater than, equal to, or less than  $\cos \pi/p$ ). The cells of this map are polyhedra  $\{n, p\}$ ,  $m$  of which meet at each edge. A sequence of vertices  $P, Q, R, S, T, \dots$  is said to form a Petrie polygon of the map if  $P, Q, R, S$  are consecutive vertices of the Petrie polygon of one cell,  $Q, R, S, T$  of another, and so on.

Let  $\{n, p, m\}_l$  denote the multiply connected map derived from  $\{n, p, m\}$  by identifying certain pairs of edges,† namely, one pair which differ by  $l$  steps along a Petrie polygon, and every pair that results from this by applying a displacement (that is, an operation of the rotation group  $[n, p, m]$ ).

The group  $((l, m, n; p))$ , involving  $l, m, n$  symmetrically, is the intrinsic rotation group of any of the six regular maps

$$\{n, p, m\}_l, \quad \{m, p, n\}_l, \quad \{l, p, n\}_m, \quad \{n, p, l\}_m, \quad \{m, p, l\}_n, \quad \{l, p, m\}_n.$$

This includes our representation of  $G^{m,n,p}$  on  $\{n, p\}_m$  as a particular case, since the three-dimensional map  $\{n, p, 2\}$  consists of just two cells, whose common boundary is the two-dimensional map  $\{n, p\}$ . The reflections of  $[n, p]$  correspond to digonal rotations of  $[n, p, 2]'$ , just as the reflections of  $[n]$  correspond to digonal rotations of  $[n, 2]'$ . The same principle was used by Todd in another connection.‡

The map  $\{4, 3, 4\}$  is of special interest, since its space is euclidean. It is, in fact, the ordinary space-filling of cubes. Its Petrie polygon is permuted by a trigonal screw.§ Therefore the derived map  $\{4, 3, 4\}_{3k}$  has  $k^3$  times as many cells as  $\{4, 3, 4\}_3$ ; that is, the order of  $((3k, 4, 4; 3))$  is  $k^3$  times that of  $((3, 4, 4; 3))$ . But we have seen that the latter, being also the rotation group of  $\{3, 3, 4\}_4$ , is of order 96. Hence  $((4, 4, 3k; 3))$  is of order  $96k^3$ .

By (4.92),  $G^{m,n,p}$  is the rotation group of any of the six maps

$$\{m, n, p\}_2, \quad \{m, p, n\}_2, \quad \{n, p, m\}_2, \quad \{n, m, p\}_2, \quad \{p, m, n\}_2, \quad \{p, n, m\}_2.$$

\* Added in proof: More precisely,  $((3, 5, l; 3))$  collapses unless  $l=3, 5, 15$ , or 30. The group is icosahedral when  $l=5$ , as well as when  $l=3$ .

† Compare the definition of  $\{m, n\}_p$  in §4.3. There we were able to say "all pairs," because we admitted reflections as well as displacements.

‡ Todd [1], p. 223.

§ Coxeter [5], p. 73.

The map  $\{m, n, p\}_2$  has one or two vertices according as  $m$  is odd or even. Reciprocally, it has one or two cells according as  $p$  is odd or even. In particular,  $\{3, 4, 3\}_2$ ,  $\{4, 3, 3\}_2$ ,  $\{3, 5, 5\}_2$ ,  $\{5, 3, 5\}_2$  each consist of a single cell. The same holds for  $\{5, 3, 3\}_3$ , since  $((3, 3, 5; 3)) \sim [5, 3]'$ . In fact,

$$\{3, 4, 3\}_2, \{5, 3, 5\}_2, \{5, 3, 3\}_3$$

are easily recognized as the *octahedron-space*\* and the two *dodecahedron-spaces*.†

TABLE I  
The known finite groups  $(l, m | n, k)$

Group	Order	$2 \sin \pi/l \sin \pi/m$ $-\cos \pi/n - \cos \pi/k$	Where group is discussed
$(2, 2   k, k) \sim [k]$	$2k$	$4 \sin^2 \pi/2k$	1.1
$(2, k   2, 2)$		$2 \sin \pi/k$	1.1
$(2i, 2j   2, 2)$	$4ij$	$2 \sin \pi/2i \sin \pi/2j$	1.1
$(4, 4   2, k)$	$4k^2$	$2 \sin^2 \pi/2k$	1.3, 2.8
$(3, 3   3, k)$		$2 \sin^2 \pi/2k$	1.6, 2.7, 2.8
$(3, k   3, 3)$	$3k^2$	$3^{1/2}(\sin \pi/k) - 1$	1.6, 1.7
$(3, 4   2, 4)$		0.518	1.5, 3.3
$(2, 4   3, 3) \sim G_{41}$	24	0.414	1.5
$(2, 3   4, 4)$		0.318	1.6
$(3, 5   2, 5)$		0.209	1.6
$(5, 5   2, 3) \sim G_{81/2} \sim \mathbb{P}_1(5)$	60	0.191	1.1, 2.5, 3.7
$(2, 5   3, 3)$		0.176	1.6
$(2, 3   5, 5)$		0.114	1.6
$(4, 6   2, 3) \sim G_{61}$	120	0.207	1.3
$(4, 5   2, 4)$	160	0.124	1.4, 2.5, 3.3
$(4, 7   2, 3)$		0.114	1.6
$(3, 3   4, 4) \sim \mathbb{P}_1(7)$	168	0.0858	1.6, 2.5, 3.5
$(3, 4   3, 4)$		0.0176	1.6, 2.7
$(4, 5   2, 5) \sim G_{61/2} \sim \mathbb{P}_1(9)$	360	0.0223	1.6
$(5, 5   2, 4)$		-0.0161	1.6, 2.5, 3.5
$(3, 3   4, 5)$		-0.0161	1.7, 2.5
$(3, 4   3, 5)$	1080	-0.0843	1.7
$(3, 5   3, 4)$		-0.189	1.7, 3.7
$(6, 7   2, 3) \sim \mathbb{P}_1(13)$	1092	-0.0661	1.7
$(7, 7   2, 3)$		-0.123	1.7, 2.5, 3.5
$(4, 8   2, 3)$	1152	+0.0412	1.3, 2.9
$(4, 9   2, 3) \sim \mathbb{P}_1(17)$	2448	-0.0163	1.7

\* Threlfall and Seifert [1], pp. 61-63.

† Weber and Seifert [1], pp. 242, 243. The hyperbolic dodecahedron-space has both two- and three-sided Petrie polygons (such as  $ab$  and  $afb$ ); the spherical dodecahedron-space has both three- and five-sided Petrie polygons (such as  $abg$  and  $figkh$ ). Since  $((3, 5, 5; 3))$  is the icosahedral group, these two spaces are defined by the symbols  $\{5, 3, 5\}_2$ ,  $\{5, 3, 3\}_2$ , just as well as by  $\{5, 3, 5\}_2$ ,  $\{5, 3, 3\}_2$ , respectively. On the other hand, since  $((3, 3, 3; 4))$  is of order 72, the map  $\{3, 4, 3\}_2$  covers  $\{3, 4, 3\}_2$  three times.

TABLE II  
The known finite groups ( $l, m, n; q$ )

Group	Order	Where group is discussed
$(bcd, cad, abd; 1) \sim [abcd]' \times [d]**$	$abcd^2$	2.2, 4.9
$(2, 2, n; n) \sim [n] \ (n \text{ odd})$	$2n$	2.3
$(2, 2, 2q; q) \sim [2q]$	$4q$	2.3, 4.9
$(2, 4, 4; q)$	$8q^2$	2.5, 2.8, 4.9
$(2, 4, 2q; 2)$	$8q^2$	2.5, 2.8, 3.4
$(2, 3, 6; q)$	$6q^2$	2.5, 2.8, 4.9
$(2, 3, 2q; 3)$	$6q^2$	2.5, 2.8, 3.4
$(3, 3, 3; q)$	$9q^2$	2.7, 2.8, 4.9
$(2, 3, 3; 2) \sim G_{31/2}$	12	2.3, 3.2, 4.9
$(2, 3, 5; 5) \sim G_{31/2}$	60	2.3, 3.2
$(2, 5, 5; 2)$	80	2.5, 4.6, 4.9
$(2, 4, 5; 3) \sim G_{31}$	120	2.5, 4.6, 4.9
$(2, 5, 6; 2) \sim G_{31}$	120	2.5, 4.6, 4.9
$(3, 3, 6; 2) \sim G_{31/2} \times G_{41/2}$	144	2.9, 4.9
$(2, 3, 7; 4) \sim \mathfrak{P}_1(7)$	168	2.3, 4.6
$(3, 4, 4; 2)$	240	2.9, 4.9
$(3, 3, 12; 2) \sim [3, 4, 3]''$	288	2.9, 4.9
$(2, 3, 8; 4) \sim \mathfrak{P}_1(7)$	336	2.5
$(3, 4, 5; 2) \sim G_{31/2}$	360	2.9, 4.9
$(3, 3, 4; 3) \sim \mathfrak{P}_1(7) \times G_3$	504	2.7, 4.9
$(2, 4, 5; 4) \sim G_{31/2} \times G_2$	720	2.5
$(2, 5, 8; 2)$	720	2.5, 3.5
$(2, 3, 7; 6) \sim \mathfrak{P}_1(13)$	1092	2.3
$(2, 3, 7; 7) \sim \mathfrak{P}_1(13)$	1092	2.3
$(2, 3, 8; 5) \sim \mathfrak{P}_1(13)$	2160	2.5, 3.5
$(2, 3, 10; 4) \sim \mathfrak{P}_1(13)$	2184	2.5
$(2, 4, 7; 3) \sim \mathfrak{P}_1(13)$	2184	2.5
$(2, 6, 7; 2) \sim \mathfrak{P}_1(13) \times G_2$	2184	2.5
$(2, 5, 9; 2) \sim \mathfrak{P}_1(19)$	3420	2.6
$(2, 3, 11; 4) \sim \mathfrak{P}_1(23)$	6072	2.6

\*  $a, b, c$  co-prime in pairs.

TABLE III  
The known finite groups  $G^{m,n,p}$

Group	Order	$\{m/n\}_p$		$1 - \cos 2\pi/m$ $-\cos 2\pi/n - \cos 2\pi/p$	Where group is discussed
		Characteristic	Genus (when orientable)		
$G^{2,n,n} \sim [n] \times G_2 \ (n \text{ even})$	$4n$	-2	0	$4 \sin^2 \pi/n$	3.2
$G^{2,n,2n} \sim [n] \times G_2 \ (n \text{ odd})$	$4n$	-2	0	$2(1 + 4 \cos^2 \pi/2n) \sin^2 \pi/2n$	3.2, 3.8
$G^{4,4,2q}$	$16q^2$	0	1	$2 \sin^2 \pi/2q$	3.4, 4.6
$G^{3,6,2q}$	$12q^2$	0	1*		
$G^{3,3,4} \sim G_{31}$	24	-2	0	2	3.2, 3.7, 4.8

\* When  $q=2$ , we have  $G^{3,4,6} \sim G_{31} \times G_2$ . The genus of  $\{3/4\}_6$  is, of course, zero.

TABLE III—Continued  
The known finite groups  $G^{m,n,p}$

Group	Order	$\{m/n\}_p$		$1 - \cos 2\pi/m$ $-\cos 2\pi/n - \cos 2\pi/p$	Where group is discussed
		Charac- teristic	Genus (when orientable)		
$G^{3,5,8} \sim G_{30/2}$	60	-1		0.882	3.2, 3.9, 4.6
$G^{3,5,10} \sim G_{30/2} \times G_2$	120	-2	0	0.392	3.2, 4.8
$G^{4,5,5} \sim (4, 5   2, 4)$	160	4			3.3, 3.8, 4.6
$G^{4,5,6} \sim G_{12} \times G_2$	240	6	4	0.191	3.5, 4.6
$G^{3,7,8} \sim \tilde{\mathfrak{P}}_1(7)$	336	4	3	0.170	3.3, 3.9, 4.6
$G^{3,7,9} \sim \mathfrak{P}_1(8)$	504	6		0.111	3.7, 4.6
$G^{3,5,5} \sim \mathfrak{P}_1(11)$	660	33		0.0729	3.7
$G^{3,8,8} \sim \tilde{\mathfrak{P}}_1(7) \times G_2$	672	14	8	0.0858	3.5
$G^{4,5,8} \sim (2, 5, 8; 2) \times G_2$	1440	36	19	-0.0161	3.5
$G^{3,7,12} \sim \tilde{\mathfrak{P}}_1(13)$	2184	26	14	+0.0105	3.3
$G^{3,7,13}$	2184	26		-0.00894	3.7, 3.9
$G^{3,7,14} \sim \tilde{\mathfrak{P}}_1(13)$	2184	26	14	-0.0245	3.3
$G^{3,9,9} \sim \mathfrak{P}_1(19)$	3420	95		-0.0321	3.7
$G^{3,8,10} \sim (2, 3, 8; 5) \times G_2$	4320	90	46	-0.0161	3.5
$G^{4,6,7} \sim \tilde{\mathfrak{P}}_1(13) \times G_2$	4368	182		-0.123	3.5
$G^{4,5,9} \sim \tilde{\mathfrak{P}}_1(19)$	6840	171		-0.0751	3.3
$G^{3,9,11} \sim \tilde{\mathfrak{P}}_1(23)$	12144	253		-0.0483	3.3
$G^{3,7,15} \sim \mathfrak{P}_1(29)$	12180	145		-0.0370	3.7

#### APPENDIX. ON THE COLLAPSE OF $(2, 3, 9; 4)$ AND $(2, 5, 7; 2)^*$

$(2, 3, 9; 4)$ : In the form (3.65), this group is defined by  $P^9 = Q^4 = (P^2Q)^3 = (P^3Q)^2 = 1$ . Since  $QP^3Q = P^{-3}$ , we have  $(P^3Q^2)^3 = 1$ . Similarly,  $^\dagger (PQP^2Q^{-1})^2 = 1$ ,  $(P^4Q)^3 = 1$ ,  $(P^2Q^2)^3 = 1$ . Using these relations, we find that

$$\begin{aligned}
 P^3Q^2P^4Q^2 &= P^7 \cdot P^{-2}Q^{-2}P^{-2} \cdot P^{-3}Q^{-1} \cdot Q^{-1} = P^7 \cdot Q^2P^2Q^2 \cdot QP^3 \cdot Q^{-1} \\
 &= P^7Q \cdot QP^2Q^{-1}P \cdot P^2Q^{-1} = P^7Q \cdot P^{-1}QP^{-2}Q^{-1} \cdot P^2Q^{-1} \\
 &= P^3(P^4QP^4)^2P^3Q \cdot Q^2P^2Q^2 \cdot Q \\
 &= P^3(Q^{-1}P^{-4}Q^{-1})^2Q^{-1}P^{-3} \cdot P^{-2}Q^{-2}P^{-2} \cdot Q \\
 &= P^3Q^{-1}P^{-2}(P^{-2}Q^{-2}P^{-2})^2P^{-3}Q^{-2}P^{-3} \cdot PQ \\
 &= P^3Q^{-1}P^{-2}(Q^2P^2Q^2)^2Q^2P^3Q^2 \cdot PQ \\
 &= P^3Q^{-1}(P^{-2}Q^{-2})^2PQ = P^3Q^{-1} \cdot Q^2P^2 \cdot PQ \\
 &= (P^3Q)^2 = 1.
 \end{aligned}$$

Hence

\* For this Appendix I am indebted to Dr. A. Sinkov. (See §2.6.)

$^\dagger$  Sinkov [3], pp. 68, 69, (2), (4), (5).

$$\begin{aligned} P^3 &= (P^{-3})^2 = (P^{-3} \cdot P^3 Q^2 P^4 Q^2)^2 \\ &= (P^2 Q^2 P^2 \cdot P^2 Q^2)^2 = (Q^{-2} P^{-2} Q^{-2} \cdot P^2 Q^2)^2 = 1. \end{aligned}$$

Thus (2, 3, 9; 4) collapses to (2, 3, 3; 2), which is the tetrahedral group.

(2, 5, 7; 2): In the form (2.12), this group is defined by  $R^5 = S^7 = (RS)^2 = (R^2 S^2)^2 = 1$ , whence

$$\begin{aligned} S^3 \cdot R^3 S R^3 S^2 R \cdot S^3 &= S \cdot S^2 R^2 \cdot R S R \cdot R^2 S^2 \cdot R S^3 = S \cdot R^{-2} S^{-2} \cdot S^{-1} \cdot S^{-2} R^{-2} \cdot R S^3 \\ &= S R \cdot R^2 S^2 R^2 \cdot R^2 S^2 \cdot S = R^{-1} S^{-1} \cdot S^{-2} \cdot S^{-2} R^{-2} \cdot S \\ &= R^2 \cdot R^2 S^2 R^2 \cdot R S = R^2 \cdot S^{-2} \cdot S^{-1} R^{-1} \\ &= R^3 \cdot R^{-1} S^{-1} \cdot S^{-2} R^{-2} \cdot R = R^3 \cdot S R \cdot R^2 S^2 \cdot R \\ &= R^3 S R^3 S^2 R. \end{aligned}$$

It follows that

$$\begin{aligned} S \cdot R^3 S R^3 S^2 R \cdot S &= R^3 S R^3 S^2 R, \quad S R^2 \cdot R S R \cdot R^2 S \cdot S R S = R^2 \cdot R S R \cdot R^2 S^2 \cdot R, \\ S R^2 \cdot S^{-1} \cdot R^2 S \cdot R^{-1} &= R^2 \cdot S^{-1} \cdot S^{-2} R^{-2} \cdot R, \end{aligned}$$

so that

$$S R^2 S^{-1} = R^2 S^{-3} R^{-1} \cdot R S^{-1} R^{-2} = R^2 S^3 R^{-2}.$$

But  $S R^2 S^{-1}$  is of period five, while  $R^2 S^3 R^{-2}$  is of period seven. Hence  $R = S = 1$ , and (2, 5, 7; 2) collapses totally.

#### REFERENCES

- Badoureaux, 1. *Mémoire sur les figures isoscèles*, Journal de l'École Polytechnique, vol. 30 (1881), pp. 47-172.
- Brahana, 1. *Regular maps and their groups*, American Journal of Mathematics, vol. 49 (1927), pp. 268-284.
- , 2. *Certain perfect groups generated by two operators of orders two and three*, American Journal of Mathematics, vol. 50 (1928), pp. 345-356.
- , 3. *Pairs of generators of the known simple groups whose orders are less than one million*, Annals of Mathematics, (2), vol. 31 (1930), pp. 529-549.
- , 4. *On the groups generated by two operators of orders two and three whose product is of order eight*, American Journal of Mathematics, vol. 53 (1931), pp. 891-901.
- Brahana and Coble, 1. *Maps of twelve countries with five sides with a group of order 120 containing an icosahedral subgroup*, American Journal of Mathematics, vol. 48 (1926), pp. 1-20.
- Burnside, 1. *Theory of Groups of Finite Order*, Cambridge, 1911.
- Carmichael, 1. *Abstract definitions of the symmetric and alternating groups and certain other permutation groups*, Quarterly Journal of Mathematics, vol. 49 (1923), pp. 226-283.
- Coxeter, 1. *The polytopes with regular-prismatic vertex figures*, Proceedings of the London Mathematical Society, (2), vol. 34 (1932), pp. 126-189.
- , 2. *Discrete groups generated by reflections*, Annals of Mathematics, (2), vol. 35 (1934), pp. 588-621.
- , 3. *Abstract groups of the form  $V_i^2 = V_i^2 = (V_i V_i)^2 = 1$* , Journal of the London Mathematical Society, vol. 9 (1934), pp. 213-219.

- , 4. *The complete enumeration of finite groups of the form  $R_4^2 = (R_i R_j)^{k_{ij}} = 1$* , Journal of the London Mathematical Society, vol. 10 (1935), pp. 21–25.
- , 5. *The groups determined by the relations  $S^l = T^m = (S^{-1}T^{-1}ST)^p = 1$* , Duke Mathematical Journal, vol. 2 (1936), pp. 61–73.
- , 6. *The abstract groups  $R^m = S^m = (R^i S^j)^{p_{ij}} = 1$ ,  $S^m = T^3 = (S^i T)^{2p_i} = 1$  and  $S^m = T^3 = (S^{-1} T S^i T)^{p_i} = 1$* , Proceedings of the London Mathematical Society, (2), vol. 41 (1936), pp. 278–301.
- , 7. *Regular skew polyhedra in three and four dimensions, and their topological analogues*, Proceedings of the London Mathematical Society, (2), vol. 43 (1937), pp. 33–62.
- Dickson, L. *Linear Groups*, Leipzig, 1901.
- Dyck, L. *Gruppentheoretische Studien*, Mathematische Annalen, vol. 20 (1882), pp. 1–44.
- Edington, L. *Abstract group definitions and applications*, these Transactions, vol. 25 (1923), pp. 193–210.
- Goursat, J. *Sur les substitutions orthogonales et les divisions régulières de l'espace*, Annales Scientifique de l'École Normale Supérieure, (3), vol. 6 (1889), pp. 9–102.
- Hamilton, W. *Memorandum respecting a new system of roots of unity*, Philosophical Magazine, (4), vol. 12 (1856), p. 446.
- Klein, F. *Vorlesungen über das Ikosaeder*, Leipzig, 1884.
- Klein-Fricke, F. *Vorlesungen über die Theorie der elliptischen Modulfunktionen*, Leipzig, 1890.
- Lewis, J. *Note on the defining relations for the simple group of order 660*, Bulletin of the American Mathematical Society, vol. 44 (1938), p. 456.
- Miller, G. *Groups generated by two operators of order three whose product is of order four*, Bulletin of the American Mathematical Society, vol. 26 (1920), pp. 361–369.
- , 2. *The commutator subgroup of a group generated by two operators*, Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 665–668.
- Schreier and van der Waerden, J. *Die Automorphismen der projektiven Gruppen*, Abhandlungen aus dem Mathematischen Seminar, Hamburg, vol. 6 (1928), pp. 203–322.
- Sinkov, E. *A set of defining relations for the simple group of order 1092*, Bulletin of the American Mathematical Society, vol. 41 (1935), pp. 237–240.
- , 2. *The groups determined by the relations  $S^l = T^m = (S^{-1}T^{-1}ST)^p = 1$* , Duke Mathematical Journal, vol. 2 (1936), pp. 74–83.
- , 3. *Necessary and sufficient conditions for generating certain simple groups by two operators of periods two and three*, American Journal of Mathematics, vol. 59 (1937), pp. 67–76.
- , 4. *Notes on the groups of genus one*, Tôhoku Mathematical Journal, vol. 43 (1937), pp. 164–170.
- , 5. *On the group-defining relations  $(2, 3, 7; p)$* , Annals of Mathematics, (2), vol. 38 (1937), pp. 577–584.
- , 6. *On generating the simple group  $LF(2, 2^N)$  by two operators of periods two and three*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 449–455.
- Threlfall, J. *Gruppenbilder*, Leipzig, 1932.
- Threlfall and Seifert, H. *Topologische Untersuchung der Diskontinuitätsbereiche endlicher Bewegungsgruppen des dreidimensionalen sphärischen Raumes*, Mathematische Annalen, vol. 104 (1931), pp. 1–70.
- Todd, J. *The groups of symmetries of the regular polytopes*, Proceedings of the Cambridge Philosophical Society, vol. 27 (1931), pp. 212–231.
- , 2. *A note on the linear fractional group*, Journal of the London Mathematical Society, vol. 7 (1932), pp. 195–200.
- Todd and Coxeter, H. *A practical method for enumerating co-sets of a finite abstract group*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 5 (1936), pp. 26–34.
- Weber, M. and Seifert, H. *Die beiden Dodekaederräume*, Mathematische Zeitschrift, vol. 37 (1933), pp. 237–253.

UNIVERSITY OF TORONTO,  
TORONTO, CANADA

# SOME EXISTENCE THEOREMS IN THE CALCULUS OF VARIATIONS

## III. EXISTENCE THEOREMS FOR NONREGULAR PROBLEMS\*

BY

E. J. McSHANE

**1. Outline of the method of proof.** This note is the third of a series, the first and second of which have already appeared in these Transactions.†

In the present paper we shall establish an existence theorem for single-integral problems of the calculus of variations without requiring quasi-regularity. In order to help the reader to find his way through the many details of this proof, we first present a heuristic outline of the method used.

Given an integral  $\mathcal{J}(C) = \int F(z, \dot{z}) dt$ , where  $z = (z^1, \dots, z^q)$ , we seek to find a rectifiable curve (in  $q$ -dimensional space) joining two fixed points  $z_1, z_2$  for which this integral is least. It is always possible to find a minimizing sequence  $\{C_n\}$ ; that is, a sequence for which  $\mathcal{J}(C_n)$  tends to the greatest lower bound  $\mu$  of  $\mathcal{J}(C)$  on the class of curves under consideration. Moreover, it is easy to find hypotheses on  $\mathcal{J}(C)$  which will ensure that a minimizing sequence exist which has a curve of accumulation  $C_0$ . That is, the curves  $C_n$  have representations  $z = z_n(t)$ ,  $(0 \leq t \leq 1)$ , for which  $z_n(t)$  tends uniformly to a limit function  $z_0(t)$ . If now we have some method of selecting the  $C_n$  so that  $z'_n(t)$  tends almost everywhere to  $z'_0(t)$ , then  $\mathcal{J}(C_n) \rightarrow \mathcal{J}(C_0)$ , and the curve  $C_0$  is the one sought.

Suppose that a curve  $C: z = z(t)$ ,  $(t_1 \leq t \leq t_2)$ , in  $q$ -dimensional space minimizes an integral  $\mathcal{J}(C)$  in the class of all curves of class‡  $D'$  joining two fixed points  $z_1$  and  $z_2$ . As is well known, the equation

$$(1.1) \quad F_i(z(t), z'(t)) = \int_{t_1}^t F_{xi}(z(t), z'(t)) + c_i$$

holds. Also, at each corner  $z(t_0)$  of  $C$  the relations

$$(1.2) \quad F_i(z(t_0), z'(t_0 - 0)) = F_i(z(t_0), z'(t_0 + 0))$$

\* Presented to the Society, December 28, 1937; received by the editors October 29, 1937 and, in revised form, February 16, 1938.

† These Transactions, vol. 44 (1938), pp. 429-438; pp. 439-453.

‡ A curve  $C$  is of class  $D'$  if it has a representation  $z = z(t)$ ,  $(t_1 \leq t \leq t_2)$ , such that the functions  $z^i(t)$  are continuous and the interval  $[t_1, t_2]$  contains a finite number of points between which  $z'(t)$  is non-vanishing and uniformly continuous. (At these points  $z'(t)$  may be  $(0, \dots, 0)$  or may fail to exist.)

and\*

$$(1.3) \quad \Omega(z(t_0), z'(t_0 - 0), z'(t_0 + 0)) \leq 0$$

hold, where

$$(1.4) \quad \Omega(z, p, r) = p^a F_{za}(z, r) - r^a F_{za}(z, p).$$

If now we choose a sequence of polygons  $\Pi_n$  joining  $z_1$  to  $z_2$  and such that  $\mathcal{Y}(\Pi_n)$  tends to the lower bound  $\mu$  of  $\mathcal{Y}(C)$  on the class of curves under consideration, we can not immediately make any statement similar to (1.1) or (1.3) for  $\Pi_n$ . However, if the number of vertices of  $\Pi_n$  is  $s_n$ , and  $\Pi_n$  actually minimizes  $\mathcal{Y}(C)$  in the class of all polygons joining  $z_1$  and  $z_2$  and having not more than  $s_n$  vertices, then it is to be expected that relations (1.1), (1.2), and (1.3) hold in an approximate sense for  $\Pi_n$ .

We now define an *approach set at  $z$*  to be an aggregate of vectors  $p$  such that

$$(1.5) \quad F_i(z, p_1) - F_i(z, p_2) = 0, \quad i = 1, \dots, q,$$

for each pair  $p_1, p_2$  of vectors of the set. In this terminology, equation (1.2) states that  $z'(t_0 - 0)$  and  $z'(t_0 + 0)$  belong to an approach set at  $z(t_0)$ . Suppose then that the polygons  $\Pi_n$  have been chosen as above so that  $\Pi_n$  minimizes  $\mathcal{Y}(C)$  in the class of polygons which have not more than  $s_n$  vertices and which join  $z_1$  to  $z_2$ . We suppose that these are represented by functions  $z = z_n(t)$ , ( $0 \leq t \leq 1$ ), which are piecewise linear on  $[0, 1]$ . Also we suppose that we have already chosen a convergent sequence, so that  $z_n(t)$  tends to a limit function  $z_0(t)$  uniformly on  $[0, 1]$ . If we fix on any number  $t_0$  in  $[0, 1]$ , it is possible to choose a subsequence  $\{\Pi_m\}$  of the sequence  $\{\Pi_n\}$  in such a way that the vectors  $z'_m(t_0 + 0)$  tend to a limit  $p_0$ . Whenever  $m$  is large and  $t$  is near  $t_0$ , the point  $z_m(t)$  is near  $z_0(t_0)$  and the vector  $z'_m(t_0 + 0)$  is near  $p_0$ . If now equation (1.1) holds, at least approximately, for  $\Pi_m$ , the value of  $F_i(z_m(t), z'_m(t))$  can change only a small amount on a short arc of  $\Pi_m$ . For by (1.1) the change in  $F_i(z_m(t), z'_m(t))$  is equal to the integral of  $F_{iz}$  over a short interval. Hence for all  $t$  near  $t_0$  the values of the  $F_i(z_m(t), z'_m(t))$  are nearly equal to the values of  $F_i(z_m(t_0), z'_m(t_0))$ , and these in turn are nearly equal to  $F_i(z_0(t_0), p_0)$ . Thus for all large  $m$  and for all  $t$  near  $t_0$  the equations

$$F_i(z_0(t_0), z'_m(t)) = F_i(z_0(t_0), p_0), \quad i = 0, 1, \dots, q,$$

are almost true, and we may expect that there is an approach set  $A$  such that  $z'_m(t)$  is in or near  $A$  whenever  $m$  is large and  $t - t_0$  small.

Suppose now that each such approach set contains only a finite number

\* Cf. I, Theorem 1.

of unit vectors, which can be set in an order  $p_1, p_2, \dots, p_k$  such that

$$(1.6) \quad \Omega(z_0(t_0), p_i, p_i) > 0, \quad j > i.$$

Now (1.3) must hold approximately on  $\Pi_m$ , and for  $t$  near  $t_0$  and  $m$  large the directions  $z'_m(t)/|z'_m(t)|$  are each near some  $p_j$ ; hence by (1.3) and (1.6) we find that a side with direction near  $p_j$  cannot precede one with direction near  $p_i$  if  $j > i$ . Thus the arc of  $\Pi_m$  near  $z_0(t_0)$  can be split into subarcs, the first of which consists of sides with directions near  $p_1$ , the second of sides with directions near  $p_2$ , and so on. That is, each such arc is almost a line segment. So is its limit arc; and for line segments  $l_n$  tending to a limit  $l_0$  it is clear that  $\mathcal{F}(l_n) \rightarrow \mathcal{F}(l_0)$ . Thus our point  $t_0$  is in an interval along which the integrand  $F(z_n, \dot{z}_n)$  converges, with arbitrarily small error, to  $F(z_0, \dot{z}_0)$ .

This argument, applied to all  $t_0$ , would yield  $\mathcal{F}(\Pi_m) \rightarrow \mathcal{F}(C_0)$ . Since  $\mathcal{F}(\Pi_m) \rightarrow \mu$ , we have  $\mathcal{F}(C_0) = \mu$ , and  $C_0$  is the curve sought.

In the following pages the argument just suggested will be generalized and made rigorous.

**2. Choice of a minimizing sequence.** We now suppose that we are given an integral  $\mathcal{F}(C)$  in parametric form, and seek the minimum of  $\mathcal{F}(C)$  in the class of all rectifiable curves  $C$  joining two given points  $z_1$  and  $z_2$ . We suppose that  $\mathcal{F}(C)$  satisfies the following condition:

(2.1) *For every constant  $M$  there is a number  $L_M$  such that all curves  $C$  joining  $z_1$  and  $z_2$  and giving  $\mathcal{F}(C)$  a value less than or equal to  $M$  have lengths not greater than  $L_M$ .*

For example, (2.1) is satisfied if there is a number  $c > 0$  such that

$$F(z, z') \geq c |z'| / (1 + |z|).$$

We shall reserve the word "vertex" for points on a polygon at which successive sides join; that is, the initial and final points will not be called vertices.

**LEMMA 1.** *Let  $\mathcal{F}(C)$  satisfy (2.1). Let  $K_s$  be the class of all polygons joining a point  $z_1$  to a point  $z_2$  and having not more than  $s$  vertices. Then for every  $s$  the class  $K_s$  contains a polygon which minimizes  $\mathcal{F}(C)$  on the class  $K_s$ .*

Let  $\mu_s$  be the greatest lower bound of  $\mathcal{F}(C)$  on the class  $K_s$ . We establish a correspondence between the polygons of  $K_s$  and the points  $(\zeta^1, \dots, \zeta^{sq})$  of  $sq$ -dimensional space by writing the coordinates of the first vertex, then those of the second, and so on.\* By (2.1), there is a number  $R$  such that if any coordinate of the point  $\zeta$  representing a polygon  $\Pi$  exceeds  $R$  in absolute value, then  $\mathcal{F}(\Pi) > \mu + 1$ . So the minimum of  $\mathcal{F}(\Pi)$  on the class  $K_s$  is to be

\* We may suppose that every polygon of  $K_s$  has  $s$  vertices, because if it has less we can insert points of division in the sides and increase the number of vertices up to  $s$ .

sought among the images of the set  $|\zeta| \leq Rsq$ . This is bounded and closed, and on it  $\mathcal{F}(\Pi)$  is continuous; hence the minimum is assumed.

Now we introduce the notation:

(2.2)  $K$  is the class of all rectifiable curves joining the two distinct fixed points  $z_1$  and  $z_2$ , and  $\mu$  is the greatest lower bound of  $\mathcal{F}(C)$  on the class  $K$ .

Then we have the following lemma:

LEMMA 2. If  $\mathcal{F}(C)$  satisfies (2.1), there exists a sequence of polygons  $\Pi_n$ :  $z = z_n(t)$ ,  $(0 \leq t \leq 1)$ , in the class  $K$  with the properties:

(a)  $\mathcal{F}(\Pi_n) \rightarrow \mu$ .

(b)  $s_n$  being the number of vertices of  $\Pi_n$ , and  $K_{s_n}$  being the subclass of  $K$  consisting of polygons of not more than  $s_n$  vertices,  $\Pi_n$  minimizes  $\mathcal{F}(C)$  on the class  $K_{s_n}$ .

(c) There is a constant  $L$  such that  $|\dot{z}_n(t)| \leq L$  for all  $n$  and all  $t$  in  $[0, 1]$ .

(d)  $z_n(t)$  converges uniformly to a limit function  $z_0(t)$ ,  $(0 \leq t \leq 1)$ .

Let  $\{C_n\}$  be a sequence of curves of  $K$  for which  $\mathcal{F}(C_n) \rightarrow \mu$ . For each fixed  $n$  we inscribe in  $C_n$  a sequence of polygons  $\{\Pi_k^*\}$  with sides tending to 0. It is well known that under these conditions  $\mathcal{F}(\Pi_k^*)$  tends to  $\mathcal{F}(C_n)$ . Hence we can find a particular  $k$  for which  $\mathcal{F}(\Pi_k^*) < \mathcal{F}(C_n) + 1/n$ .

Let  $s_k^*$  be the number of vertices of  $\Pi_k^*$ . By Lemma 1, there is a polygon  $\Pi_n$  of  $s_n$  vertices, where  $s_n \leq s_k^*$ , which minimizes  $\mathcal{F}(C)$  on the class  $K_{s_n}$ . Therefore  $\mathcal{F}(\Pi_n) \leq \mathcal{F}(\Pi_k^*) < \mathcal{F}(C_n) + 1/n$ . Since  $\Pi_n$  minimizes  $\mathcal{F}(C)$  in the class  $K_{s_n}$  and is in the class  $K_{s_n} \subset K_{s_k^*}$ , it is clear that  $\Pi_n$  minimizes  $\mathcal{F}(C)$  on  $K_{s_n}$ . Also,

$$\limsup_{n \rightarrow \infty} \mathcal{F}(\Pi_n) \leq \lim_{n \rightarrow \infty} \mathcal{F}(C_n) = \mu.$$

But  $\Pi_n$  is in  $K$ ; so  $\mathcal{F}(\Pi_n) \geq \mu$ , and  $\liminf_{n \rightarrow \infty} \mathcal{F}(\Pi_n) \geq \mu$ . Thus (a) and (b) are established.

There is no loss of generality in supposing  $\mathcal{F}(\Pi_n) < \mu + 1$  for all  $n$ . Hence, by hypothesis (2.1), there is an  $L$  such that each  $\Pi_n$  has length  $\mathcal{L}(\Pi_n)$  less than  $L$ . On  $\Pi_n$  we introduce the parameter  $t = s/\mathcal{L}(\Pi_n)$ , where  $s$  is arc length. Then  $\Pi_n$  has a representation  $z = z_n(t)$ ,  $(0 \leq t \leq 1)$ , with  $|\dot{z}_n(t)| \leq \mathcal{L}(C_n) < L$ , and (c) is satisfied.

Finally, the  $z_n(t)$  all satisfy a Lipschitz condition with constant  $L$ ; so by Ascoli's theorem there is a subsequence (which we may without loss of generality suppose to be the whole sequence) for which  $z_n(t)$  converges uniformly to a limit function  $z_0(t)$ . This completes the proof of the lemma.

**Remark.** If a sequence  $\{\Pi_n\}$  satisfies (a), (b), (c), and (d) of Lemma 2, so does every subsequence of  $\{\Pi_n\}$ .

Before stating the next lemma we introduce a definition.

(2.3) If  $[\alpha, \beta]$  is an interval having points in common with  $[0, 1]$  and  $t_{n,1}, t_{n,2}, \dots, t_{n,k}$  are (in that order) the values of  $t$  in  $[\alpha, \beta]$  which define vertices of  $\Pi_n$ , then

$$\theta(n, \alpha, \beta) \equiv t_{n,1}, \quad \Theta(n, \alpha, \beta) \equiv t_{n,k}.$$

If no value of  $t$  in  $[\alpha, \beta]$  defines a vertex of  $\Pi_n$ , we define

$$(2.4) \quad \theta(n, \alpha, \beta) = \Theta(n, \alpha, \beta) = (\alpha + \beta)/2.$$

Then we have the following lemma:

LEMMA 3. If  $\{\Pi_n\}$  is the sequence of Lemma 2, then for every  $\epsilon > 0$  and every  $t_0$  in  $[0, 1]$  there is a  $\delta > 0$  such that

$$(2.5) \quad |F_i(z_0(t_0), z_n'(t')) - F_i(z_0(t_0), z_n'(t''))| \leq \epsilon, \quad i = 1, \dots, q,$$

if  $t_0, t'$ , and  $t''$  are all in the (open) interval  $(\theta(n, t_0 - \delta, t_0 + \delta), \Theta(n, t_0 - \delta, t_0 + \delta))$  and  $z_n(t')$  and  $z_n(t'')$  are not vertices† of  $\Pi_n$ .

We shall suppose, to be specific, that  $t' < t''$ . Let  $t_1, t_2, \dots, t_k$  define successive vertices of  $\Pi_n$  such that  $t_1 < t' < t_2$  and  $t_{k-1} < t'' < t_k$ . By the definition of  $\theta$  and  $\Theta$  we know that

$$(2.6) \quad t_0 - \delta < t_1 < t_k < t_0 + \delta.$$

From  $\Pi_n$  we now form the polygon  $\Pi_n^j(\tau)$  by displacing each of the vertices  $z(t_1), \dots, z(t_{k-1})$  by the amount  $(0, \dots, \tau, 0, \dots, 0)$ , the  $\tau$  being in the  $j$ th place. Because of the minimizing property of  $\Pi_n$  we have

$$(2.7) \quad \left. \frac{d}{d\tau} \mathcal{Y}(\Pi_n^j(\tau)) \right|_{\tau=0} = 0.$$

But if we apply formula (2.8) of I to the successive sides of  $\Pi_n^j(\tau)$  defined by  $t_1 \leq t \leq t_2, \dots, t_{k-1} \leq t \leq t_k$ , and then add, we obtain

$$(2.8) \quad \begin{aligned} \left. \frac{d}{d\tau} \mathcal{Y}(\Pi_n^j(\tau)) \right|_{\tau=0} &= F_j(z_n(\bar{t}), z_n'(\bar{t})) - F_j(z_n(\bar{t}), z_n'(\bar{t})) \\ &+ \int_{t_1}^{t_2} F_{zj}(z_n, \dot{z}_n) \frac{t - t_1}{t_2 - t_1} dt + \int_{t_2}^{t_{k-1}} F_{zj}(z_n, \dot{z}_n) dt \\ &+ \int_{t_{k-1}}^{t_k} F_{zj}(z_n, \dot{z}_n) \frac{t_k - t}{t_k - t_{k-1}} dt, \end{aligned}$$

† It is easy to see that this restriction is essentially no restriction, for if  $t'$  defines a vertex, we first write (2.5) with  $t'$  replaced by  $t' + h$ ,  $h > 0$ , and then let  $h \rightarrow 0$ . For small  $h$  the value of  $z'(t' + h)$  is constantly equal to  $z'(t' + 0)$ ; whence we find (2.5) with  $t'$  replaced by  $t' + 0$ . Similarly, we could replace  $t'$  by  $t' - 0$ , and analogously for  $t''$ .

where  $t_1 < \bar{t} < t_2$  and  $t_{k-1} < \bar{t} < t_k$ . All the curves  $\Pi_n$  lie in a bounded closed subset of the space, and all  $|\dot{z}_n|$  are less than  $L$ . Hence the functions  $F_{zi}(z_n, \dot{z}_n)$  are bounded, say less than  $N$  in absolute value. Then the sum of the three integrals on the right is at most  $N(t_k - t_1)$  in absolute value, so that

$$(2.9) \quad 0 = F_i(z_n(\bar{t}), z'_n(\bar{t})) - F_i(z_n(\bar{t}), z'_n(\bar{t})) + \theta_i,$$

where  $|\theta_i| \leq N(t_k - t_{k-1}) \leq 2N\delta$ .

On the interval  $(t_1, t_2)$  the derivative  $z'_n(t)$  is constantly equal to  $z'_n(t')$ , and on  $(t_{k-1}, t_k)$  it is constantly equal to  $z'_n(t'')$ . Hence from (2.9) we obtain

$$(2.10) \quad \begin{aligned} & |F_i(z_n(t'), z'_n(t')) - F_i(z_n(t''), z'_n(t''))| \\ & \leq |F_i(z_n(t'), z'_n(t')) - F_i(z_n(\bar{t}), z'_n(t'))| \\ & \quad + |F_i(z_n(t''), z'_n(t'')) - F_i(z_n(\bar{t}), z'_n(t''))| + |\theta_i|. \end{aligned}$$

Because of the continuity of  $F_i$ , each of the first two terms on the right is less than  $\epsilon/3$  if  $|z_n(\bar{t}) - z_n(t')|$  and  $|z_n(\bar{t}) - z_n(t'')|$  are less than a certain  $\gamma > 0$ . This is surely the case if  $|\bar{t} - t'|$  and  $|\bar{t} - t''|$  are less than  $\gamma/L$ , which again is surely true if  $2\delta < \gamma/L$ . Also, if  $\delta < \epsilon/6N$ , the term  $|\theta_i|$  is at most  $2N\delta < \epsilon/3$ . Hence, if  $\delta$  is any number less than the smaller of  $\gamma/2L$  and  $\epsilon/6N$ , inequality (2.10) implies (2.5).

**Remark.** From the last sentence of the proof it is obvious that  $\delta$  can be chosen as near to 0 as desired.

**3. Directions of the sides of the minimizing sequence.** The next lemma will explain the name "approach set" given to the sets satisfying (1.5).

**LEMMA 4.** Let  $\{\Pi_n\}$  be a sequence of polygons satisfying (a), (b), (c), and (d) of Lemma 2, and let  $z_1 \neq z_2$ . Let  $t_0$  be interior to  $(0, 1)$ . Then there is a subsequence  $\{\Pi_m\}$  of  $\{\Pi_n\}$  and an approach set  $A$  at  $z_0(t_0)$  such that for every  $\epsilon > 0$  there is a  $\delta > 0$  and an  $m_0$  such that  $z'_m(t)$  is in the  $\epsilon$ -neighborhood of  $A$  whenever  $z_m(t)$  is not a vertex,  $m > m_0$ , and

$$(3.1) \quad \theta(m, t_0 - \delta, t_0 + \delta) < t < \Theta(m, t_0 - \delta, t_0 + \delta).$$

If there is a  $\delta > 0$  and a subsequence  $\{\Pi_m\}$  such that not more than one  $t$  in  $(t_0 - \delta, t_0 + \delta)$  defines a vertex of  $\Pi_m$ , then the conclusion holds vacuously, for then  $\theta(m, t_0 - \delta, t_0 + \delta) = \Theta(m, t_0 - \delta, t_0 + \delta)$ , and (3.1) is not satisfied by any  $t$ .

Otherwise, for each integer  $m$  there are infinitely many  $n_m$  such that the interval

$$(3.2) \quad (\theta(n_m, t_0 - m^{-1}, t_0 + m^{-1}), \Theta(n_m, t_0 - m^{-1}, t_0 + m^{-1}))$$

has length greater than zero. We can suppose  $n_m$  so chosen that  $n_1 < n_2 < \dots$

In each interval (3.2) we choose a number  $t_{n_m}$  which does not define a vertex of  $\Pi_{n_m}$ . The vectors  $z'_{n_m}(t_{n_m})$  all have lengths less than or equal to  $L$ ; hence they have an accumulation vector  $p_0$ . We select a subsequence  $\{m\}$  of the sequence  $\{n_m\}$  such that  $z'_m(t_m) \rightarrow p_0$ .

Now let  $A$  be the approach set at  $z_0(t_0)$  which contains  $p_0$ , and let  $U$  be the set of all vectors  $u$  whose distance from  $A$  is less than  $\epsilon$ . Then  $U$  is clearly open. We are to show that  $z'_m(t)$  is in  $U$  if  $t$  and  $m$  satisfy the conditions of our lemma. If  $V$  is the set of all vectors  $v$  of length  $|v| \leq L$  which are not in  $U$ , then  $V$  is bounded and closed and has no point in common with  $A$ . Hence the function

$$(3.3) \quad \sum_{i=1}^q |F_i(z_0(t_0), p_0) - F_i(z_0(t_0), v)|$$

is positive for all  $v$  in  $V$ . Being continuous<sup>†</sup> on  $V$ , it is greater than a positive number  $3\gamma$ .

Now

$$(3.4) \quad \begin{aligned} & \sum_{i=1}^q |F_i(z_0(t_0), z'_m(t)) - F_i(z_0(t_0), p_0)| \\ & \leq \sum_{i=1}^q |F_i(z_0(t_0), z'_m(t)) - F_i(z_m(t), z'_m(t))| \\ & \quad + \sum_{i=1}^q |F_i(z_m(t), z'_m(t)) - F_i(z_m(t_m), z'_m(t_m))| \\ & \quad + \sum_{i=1}^q |F_i(z_m(t_m), z'_m(t_m)) - F_i(z_0(t_0), p_0)|. \end{aligned}$$

Here, as in Lemma 3, the parameter  $t$  is  $s/\mathcal{L}(\Pi_m)$ , so that  $|z'_m(t)| = \mathcal{L}(\Pi_m) \geq |z_2 - z_1| > 0$ . Thus the arguments in the first term on the right have a bounded closed range and  $z'_m(t)$  is bounded from 0. Hence the first term is less than  $\gamma$  if  $|z_0(t_0) - z_m(t)|$  is smaller than a certain number  $\kappa$ . Since  $z_m(t)$  tends uniformly to  $z_0(t)$  and  $z_0(t)$  is continuous, this is true if  $\delta$  is small enough and  $m$  is greater than a certain  $m_0$ .

In the last term on the right,  $z'_m(t_m)$  tends to  $p_0$  and  $z_m(t_m)$  to  $z_0(t_0)$ ; so this term is less than  $\gamma$  if  $m$  is greater than a certain  $m_2$ .

If  $\delta$  is small enough and  $m$  is greater than a certain  $m_3$ , then by Lemma 3 the third term is less than  $\gamma$  whenever  $t$  does not define a vertex of  $\Pi_m$  and (3.1) holds. Hence if  $\delta$  is small enough and  $m > m_0 \equiv \max(m_1, m_2, m_3)$ , then for all  $t$  satisfying the requirements of our lemma we have

<sup>†</sup> The only point at which  $F_i(z_0(t_0), v)$  is discontinuous is  $v=0$ , which is a limit point of  $A$  and therefore not in  $V$ .

$$\sum_{i=1}^q |F_i(z_0(t_0), z_m'(t)) - F_i(z_0(t_0), p_0)| < 3\gamma.$$

But if  $z_m'(t)$  were in  $V$ , this expression (see (3.3)) would exceed  $3\gamma$ . So  $z_m'(t)$  is not in  $V$ . Also, it satisfies the condition  $|z_m'(t)| \leq L$ . Hence it must belong to  $U$ , and the lemma is established.

**Remark.** Again it is clear that the  $\delta$  can be chosen as small as desired.

**4. Proof of the principal theorem.** We now impose a new hypothesis on the integral  $\mathcal{J}(C)$ . We shall make the following supposition:

(4.1) *For every  $z$ , each approach set  $A$  at  $z$  is the sum of a finite number of convex sets  $A_1, \dots, A_k$ , which can be so ordered that*

$$\Omega(z, p_i, p_j) < 0$$

*if  $p_i$  is in  $A_i$  and  $p_j$  in  $A_j$ , ( $i < j$ ).*

Our principal theorem is the following:

**THEOREM 1.** *If  $F(z, z')$  satisfies conditions (2.1) and (4.1), then in the class  $K$  of all rectifiable curves joining two distinct points  $z_1$  and  $z_2$  there is a curve which minimizes the integral*

$$\mathcal{J}(C) = \int_C F(z, \dot{z}) dt.$$

We select a sequence of polygons  $\Pi_n$  and a curve  $C_0$  satisfying the conclusions of Lemma 2, and we define

$$(4.2) \quad \phi_n(t) = \int_0^t F(z_n, \dot{z}_n) dt, \quad 0 \leq t \leq 1; n = 0, 1, 2, \dots$$

Since  $|\dot{z}_n(t)| \leq L$ , the integrands are bounded, and the  $\phi_n$  all satisfy the same Lipschitz condition. Hence by Ascoli's theorem we can select a subsequence (we suppose it to be the whole sequence) such that  $\phi_n(t)$  tends uniformly to a limit function  $\phi(t)$ . By (a) of Lemma 2,

$$(4.3) \quad \phi(1) = \lim_{n \rightarrow \infty} \phi_n(1) = \lim_{n \rightarrow \infty} \mathcal{J}(\Pi_n) = \mu.$$

Since

$$(4.4) \quad \phi_0(1) = \mathcal{J}(C_0),$$

we must show then that

$$(4.5) \quad \phi_0(1) = \phi(1) = \mu.$$

We shall in fact show that

$$(4.6) \quad \phi_0(t) = \phi(t), \quad 0 \leq t \leq 1.$$

Let  $t_0$  be a point interior to  $(0, 1)$  at which  $\phi'(t)$  and  $\phi_0'(t)$  are both defined. We shall first show

$$(4.7) \quad \phi'(t_0) = \phi_0'(t_0).$$

This will be established if we can prove the following statement:

(4.8) *For every positive number  $\gamma$  there is a  $t^* > t_0$  such that if  $t_0 < \rho < \sigma < t^*$ , then*

$$(4.9) \quad |[\phi(\sigma) - \phi(\rho)] - [\phi_0(\sigma) - \phi_0(\rho)]| \leq \gamma(\sigma - \rho).$$

For if (4.8) holds, then by continuity (4.9) holds with  $\rho$  replaced by  $t_0$ . Dividing by  $\sigma - t_0$  and letting  $\sigma$  tend to  $t_0$  we obtain

$$|\phi'(t_0) - \phi_0'(t_0)| \leq \gamma.$$

Since  $\gamma$  is an arbitrary positive number, this can be true only if (4.7) holds. We therefore take  $\gamma$  to be a positive number and proceed to establish (4.8).

Let us first dispose of the relatively simple case in which there is a subsequence  $\{\Pi_h\}$  and a  $\delta > 0$  such that either (a)  $\lim_{h \rightarrow \infty} \theta(h, t_0 - \delta, t_0 + \delta) > t_0$  or (b)  $\lim_{h \rightarrow \infty} \Theta(h, t_0 - \delta, t_0 + \delta) \leq t_0$ . If (a) holds, we denote  $\lim \theta$  by  $t^*$ ; if (b) holds, we let  $t^*$  equal  $t_0 + \delta$ . Let  $\rho, \sigma$  be numbers such that  $t_0 < \rho < \sigma < t^*$ . Then if  $h$  is large enough, the interval  $[\rho, \sigma]$  is contained in case (a) in  $[t_0 - \delta, \theta(h, t_0 - \delta, t_0 + \delta)]$ , and in case (b) in  $[\Theta(h, t_0 - \delta, t_0 + \delta), t_0 + \delta]$ . In either case,  $z_h(t)$  is linear on  $[\rho, \sigma]$  for all large  $h$ . The same is then true of its limit  $z_0(t)$ , and it follows at once that  $z_h'(t)$  tends to  $z_0'(t)$  uniformly on  $[\rho, \sigma]$ . Therefore

$$\begin{aligned} \phi(\sigma) - \phi(\rho) &= \lim_{h \rightarrow \infty} \{\phi_h(\sigma) - \phi_h(\rho)\} = \lim_{h \rightarrow \infty} \int_{\rho}^{\sigma} F(z_h, z_h') dt \\ &= \int_{\rho}^{\sigma} F(z_0, z_0') dt = \phi_0(\sigma) - \phi_0(\rho); \end{aligned}$$

and (4.8) is established for all positive  $\gamma$  at once.

Now we return to the general case. By Lemma 4, there is an approach set  $A$  at  $z_0(t_0)$  with the properties there specified. If  $p_0$  is any fixed vector in  $A$ , then for every vector  $p$  in  $A$  the equations

$$F_i(z_0(t_0), p) = F_i(z_0(t_0), p_0), \quad i = 1, \dots, q,$$

hold, by definition. If we denote the right-hand members of these equations by  $l_1, \dots, l_q$ , respectively, then on multiplying by  $p^i$  and summing we find, for all  $p$  in  $A$ , that

$$F(z_0(t_0), p) = l_{\alpha} p^{\alpha}.$$

The set of vectors  $p$  in  $A$  with  $|p| \leq L$  is bounded and closed, and the function  $F$  is continuous. Therefore there is a  $\lambda > 0$  such that

$$(4.10) \quad |F(z, p) - l_\alpha p^\alpha| < \gamma/2$$

if  $|z - z_0(t_0)| < \lambda$  and  $p$  is in the  $\lambda$ -neighborhood  $U_\lambda$  of the intersection of  $A$  with the sphere  $|p| \leq L$ .

The set  $A$  is the sum of convex sets  $A_1, \dots, A_k$ . Let us denote by  $U_{j,\epsilon}$  the  $\epsilon$ -neighborhood of the intersection of  $A_j$  with the sphere  $|p| \leq L$ . Then  $U_\epsilon = U_{1,\epsilon} + \dots + U_{k,\epsilon}$ . By hypothesis, the  $A_j$  have the property that

$$(4.11) \quad \Omega(z_0(t_0), p_i, p_j) < 0$$

if  $p_i$  is in  $A_i$ ,  $p_j$  is in  $A_j$ , and  $i < j$ . Therefore on the bounded closed set of arguments  $(z, p_i, p_j)$  which satisfy the conditions  $z = z_0(t_0)$ ,  $p_i$  in  $A_i$ ,  $p_j$  in  $A_j$ , ( $i < j$ ),  $|z_2 - z_1| \leq |p_i| \leq L$ ,  $|z_2 - z_1| \leq |p_j| \leq L$  this function has a negative upper bound  $-2\kappa$ . Since it is continuous, we find that there is an  $\eta > 0$  such that

$$(4.12) \quad \Omega(z, p_i, p_j) < -\kappa$$

whenever  $p_i$  is in  $U_{i,\eta}$ ,  $p_j$  is in  $U_{j,\eta}$ , ( $i < j$ ),  $|z_2 - z_1| \leq |p_i| \leq L$ ,  $|z_2 - z_1| \leq |p_j| \leq L$ , and  $|z - z_0(t_0)| < 3\eta$ .

Now define

$$(4.13) \quad \epsilon = \min(\eta, \lambda).$$

We use this for the  $\epsilon$  in Lemma 4, and choose (see the remark after the lemma) a value for  $\delta$  which is less than  $\epsilon/2(L+1)$ . Furthermore, from the sequence  $\{\Pi_m\}$  of that lemma we discard those polygons (finite in number) for which the inequality

$$(4.14) \quad |z_m(t) - z_0(t)| < \epsilon/2, \quad 0 \leq t \leq 1,$$

fails to hold. Then, by Lemma 4, together with (4.10) and (4.13), we find:

(4.15) For all functions  $z_m(t)$  and all  $t$  such that

$$\theta(m, t_0 - \delta, t_0 + \delta) < t < \Theta(m, t_0 - \delta, t_0 + \delta),$$

the inequality

$$|F(z_m(t), \dot{z}_m(t)) - l_\alpha \dot{z}_m^\alpha(t)| < \gamma/2$$

holds, and  $z_m'(t)$  is in one of the sets  $U_{1,\epsilon}, \dots, U_{k,\epsilon}$  if  $z_m'(t)$  is defined.

Let  $t_1, t_2, t_3$  define three consecutive vertices of  $\Pi_m$ , and let the inequalities

$$\theta(m, t_0 - \delta, t_0 + \delta) \leq t_1 < t_2 < t_3 \leq \Theta(m, t_0 - \delta, t_0 + \delta)$$

hold. The sides of  $\Pi_m$  corresponding to  $[t_1, t_2]$  and  $[t_2, t_3]$  have directions

$z_m'(t_1+0)$ ,  $z_m'(t_2+0)$ . By (4.15) these belong respectively to neighborhoods  $U_{i,\epsilon}$ ,  $U_{j,\epsilon}$ ; and the inequalities

$$|z_2 - z_1| \leq |z_m'(t)| \leq L$$

always hold. We now show that  $i$  is less than  $j$ . Suppose the contrary. If we interchange these two sides of  $\Pi_m$ , we obtain a new polygon  $\Pi_m^*$ , and by formula (3.19) of I

$$(4.16) \quad \mathcal{Y}(\Pi_m^*) - \mathcal{Y}(\Pi_m) = -|z_m(t_3) - z_m(t_2)| |z_m(t_2) - z_m(t_1)| \cdot \Omega(\bar{z}, z_m'(t_1+0), z_m'(t_2+0)),$$

where  $\bar{z}$  is in the parallelogram determined by the two sides. However,  $|t_1 - t_0| < \delta < \epsilon/2L$ ; so  $|z_m(t_1) - z_m(t_0)| < L \cdot \epsilon/2L = \epsilon/2$ , while by (4.14),  $|z_m(t_0) - z_0(t_0)| < \epsilon/2$ . Hence  $|z_m(t_1) - z_0(t_0)| < \epsilon$ . The sides of this parallelogram have lengths less than  $|t_3 - t_1| L < 2\delta L < \epsilon$ . Thus the whole parallelogram lies in the  $3\epsilon$ -neighborhood of  $z_0(t_0)$ , and in particular  $|\bar{z} - z_0(t_0)| < 3\epsilon \leq 3\eta$ , by (4.13). Therefore by (4.12) and (4.13) the factor  $\Omega$  in (4.16) is positive,† and  $\mathcal{Y}(\Pi_m^*) < \mathcal{Y}(\Pi_m)$ . This contradicts the minimizing property of  $\Pi_m$ , and establishes the inequality  $i < j$ .

We have therefore established, as a sort of addition to (4.15), that for  $t$  between  $\theta$  and  $\Theta$  no side of  $\Pi_m$  whose direction is in a  $U_{j,\epsilon}$  is followed by one whose direction is in a  $U_{i,\epsilon}$  with  $i < j$ . Let us now define

$$(4.17) \quad t_{m,0} = \theta(m, t_0 - \delta, t_0 + \delta), \quad t_{m,k} = \Theta(m, t_0 - \delta, t_0 + \delta).$$

By the previous remark, there are numbers  $t_{m,i}$  defining vertices of  $\Pi_m$  such that

$$t_{m,0} \leq t_{m,1} \leq \dots \leq t_{m,k},$$

and if  $t_{m,j-1} < t < t_{m,j}$  and  $z_m'(t)$  is defined, then  $z_m'(t)$  is in  $U_{j,\epsilon}$ . We now select a subsequence  $\{\Pi_h\}$  of  $\{\Pi_m\}$  such that for each  $j=0, 1, \dots, k$  the limit

$$\tau_j = \lim_{h \rightarrow \infty} t_{h,j}$$

exists. We may suppose  $\tau_0 \leq t_0 < \tau_k$ , for otherwise we are back to the simpler case first considered.

There is a first  $j$  such that  $\tau_j > t_0$ ; hence  $\tau_{j-1} \leq t_0 < \tau_j$ . Let  $\rho, \sigma$  be numbers such that  $t_0 < \rho < \sigma < \tau_j$ . For all large  $h$  the inequalities  $\tau_{h,j-1} < \rho < \sigma < \tau_{h,j}$  hold, so that  $\dot{z}_h(t)$  is in  $U_{j,\epsilon}$  if  $\rho \leq t \leq \sigma$ . The set  $U_{j,\epsilon}$  is the  $\epsilon$ -neighborhood of the

† Recall that the interchange of  $p_i$  and  $p_j$  in (4.12) changes the sign of  $\Omega$ .

intersection of two convex sets,  $A_j$  and the sphere  $|p| \leq L$ ; so  $U_{j,\epsilon}$  is convex. By Jensen's inequality (in geometric form), if  $\rho \leq t < t' \leq \sigma$ , the vector

$$\frac{1}{t' - t} \int_t^{t'} \dot{z}_h(t) dt = \frac{z_h(t') - z_h(t)}{t' - t}$$

is in the closure  $\bar{U}_{j,\epsilon}$ . Letting  $h \rightarrow \infty$ , we find that the limit

$$\frac{z_0(t') - z_0(t)}{t' - t}$$

is also in  $\bar{U}_{j,\epsilon}$ . Now let  $t' \rightarrow t$ ; we find that the vector  $z'_0(t)$  is in  $\bar{U}_{j,\epsilon}$  if it is defined. Since the zero vector is in  $U_{j,\epsilon}$ , we see that in any case  $\dot{z}_0(t)$  is in  $\bar{U}_{j,\epsilon}$  for  $\rho \leq t < \sigma$ .

Turning now to (4.10), we therefore obtain

$$|F(z_0(t), \dot{z}_0(t)) - l_\alpha \dot{z}_0^\alpha(t)| \leq \gamma/2$$

if  $\rho \leq t < \sigma$ . Hence

$$(4.18) \quad \left| \phi_0(\sigma) - \phi_0(\rho) - l_\alpha \{z_0^\alpha(\sigma) - z_0^\alpha(\rho)\} \right| = \left| \int_\rho^\sigma \{F(z_0, \dot{z}_0) - l_\alpha \dot{z}_0^\alpha\} dt \right| \\ \leq \gamma(\sigma - \rho)/2.$$

On the other hand, by (4.15)

$$(4.19) \quad \left| \phi_h(\sigma) - \phi_h(\rho) - l_\alpha \{z_h^\alpha(\sigma) - z_h^\alpha(\rho)\} \right| = \left| \int_\rho^\sigma \{F(z_h, \dot{z}_h) - l_\alpha \dot{z}_h^\alpha\} dt \right| \\ < \gamma(\sigma - \rho)/2.$$

If we here let  $h \rightarrow \infty$ , we obtain

$$(4.20) \quad \left| \phi(\sigma) - \phi(\rho) - l_\alpha \{z_0^\alpha(\sigma) - z_0^\alpha(\rho)\} \right| \leq \gamma(\sigma - \rho)/2.$$

From (4.18) and (4.20) we at once obtain (4.9). Thus statement (4.8), and with it (4.7), is established.

Since  $\phi(0) = \phi_0(0) = 0$  and (4.7) holds for almost all  $t$ , and since the functions  $\phi$  and  $\phi_0$  are both absolutely continuous, it follows that  $\phi(t) \equiv \phi_0(t)$ , so that (4.5) is established. Therefore the curve  $z = z_0(t)$  is the minimizing curve sought.

**5. Example.** If  $F(z, z')$  satisfies (2.1) and is quasi-regular† (see (6.2)), every approach set  $A$  at every point  $z_0$  consists of a single convex set; so (4.1) is trivially satisfied. Hence for such integrands there is always a minimizing

† Necessarily positive quasi-regular, for if it is negative quasi-regular and not linear, (2.1) cannot hold.

curve for  $\mathcal{Y}(C)$  in the class  $K$  of all curves joining two given points  $z_1$  and  $z_2$ . This is, of course, a special case of a known theorem.

An example of an essentially different type is  $\int F dt$  where

$$F(x, y, x', y') = (x'^2 + y'^2)^{1/2} + y^2[4(x'^2 + y'^2)^{1/2} - (x'^2 + 8y'^2)^{1/2}].$$

This integrand is not positive quasi-regular, for if  $|y| > 1/2$ , the  $\mathcal{E}$ -function can be negative. Given a set  $(x, y, p, q)$  with  $p^2 + q^2 = 1$ , we seek to determine all  $(\bar{p}, \bar{q})$  with  $\bar{p}^2 + \bar{q}^2 = 1$  such that

$$(5.1) \quad F_{x'}(x, y, \bar{p}, \bar{q}) = F_{x'}(x, y, p, q), \quad F_{y'}(x, y, \bar{p}, \bar{q}) = F_{y'}(x, y, p, q).$$

We can write (5.1) in the form

$$(5.2) \quad \begin{aligned} p[1 + 4y^2 - y^2(8 - 7p^2)^{-1/2}] &= \bar{p}[1 + 4y^2 - y^2(8 - 7\bar{p}^2)^{-1/2}], \\ q[1 + 4y^2 - 8y^2(1 + 7q^2)^{-1/2}] &= \bar{q}[1 + 4y^2 - 8y^2(1 + 7\bar{q}^2)^{-1/2}]. \end{aligned}$$

The function in the left-hand member of the first of these equations is easily seen to have a positive derivative with respect to  $p$  for  $|p| \leq 1$ ; so the only solution of the first equation is  $\bar{p} = p$ . Since  $p^2 + q^2 = \bar{p}^2 + \bar{q}^2 = 1$ , this implies  $\bar{q} = \pm q$ . Hence we substitute  $-q$  for  $\bar{q}$  in the second equation of (5.2). If it is not satisfied, then the entire approach set at  $(x, y)$  containing  $(p, q)$  is the set of multiples  $(kp, kq)$ , ( $k > 0$ ). If it is satisfied, then

$$1 + 4y^2 - 8y^2(1 + 7q^2)^{-1/2} = 0$$

or

$$(5.3) \quad q = \pm \left[ \frac{1}{7} \left\{ \left( \frac{8y^2}{1 + 4y^2} \right)^2 - 1 \right\} \right]^{1/2}.$$

If  $q$  has the value (5.3), then  $A$  consists of the positive multiples of  $(p, q)$  and of  $(p, -q)$ , which are distinct if  $q \neq 0$ . Equation (5.3) can hold with  $q \neq 0$  only for  $|y| > 1/2$ , and it is easy to see that  $|q| < (3/7)^{1/2}$ .

In any case, the approach set at  $(x, y)$  containing  $(p, q)$  consists either of a half-line or of two half-lines; and a half-line is a convex set. If the set consists of a single half-line, (4.1) clearly holds. Otherwise, let  $A$  consist of the multiples of  $(p, q)$  and of  $(p, -q)$ . If  $y > 0$ , we define  $A_1$  to be the multiples of  $(p, -|q|)$  and  $A_2$  to be the multiples of  $(p, |q|)$ ; if  $y < 0$ , we interchange the definitions of  $A_1$  and  $A_2$ . Then each vector in  $A_1$  has the form  $(p_1, q_1) = (k_1 p, -k_1 \operatorname{sgn} y |q|)$ , ( $k_1 > 0$ ), and each vector in  $A_2$  has the form  $(p_2, q_2) = (k_2 p, k_2 \operatorname{sgn} y |q|)$ , ( $k_2 > 0$ ). Hence

$$\begin{aligned} \Omega_F(x, y; p_1, q_1; p_2, q_2) &= 2y \{ q_1 [4(p_2^2 + q_2^2)^{1/2} - (p_2^2 + 8q_2^2)^{1/2}] \\ &\quad - q_2 [4(p_1^2 + q_1^2)^{1/2} + (p_1^2 + 8q_1^2)^{1/2}] \}. \end{aligned}$$

The factors in square brackets are positive, and  $yq_1$  and  $-yq_2$  are negative ( $q_1$  and  $q_2$  are not zero, since by hypothesis  $A$  contains more than one unit vector). Hence the left member is negative, and (4.1) holds.

6. **Generalization to sets  $S$  with boundary points.** A slight generalization of Theorem 1 can be established at once. We need not assume that  $z_2 \neq z_1$ . The only use made of this hypothesis was to ensure that  $|z'_n(t)|$  had a positive lower bound. If  $z_2 = z_1$ , we distinguish two cases.

**Case I.** The  $|z'_n(t)|$  have a positive lower bound. In this case the preceding proof applies without change.

**Case II.** The  $\liminf |z'_n(t)| = 0$ . Since  $|z'_n(t)| = \mathcal{L}(\Pi_n)$  for almost all  $t$ , this implies that for a subsequence  $\{\Pi_m\}$  of the minimizing sequence  $\{\Pi_n\}$  we have  $\mathcal{L}(\Pi_m) = 0$ . Then  $\mu = \lim \mathcal{Y}(\Pi_n) = 0$ ; so the degenerate curve consisting of the single point  $z_1$  is the curve sought.

Less trivial is the generalization which allows us to study integrands not defined for all  $z$ . Before stating this theorem it is desirable to introduce some definitions and establish a lemma.

(6.1) *The function  $F(z, z')$  is positive (negative) regular the point  $z_0$  if†  $u^a F_{\alpha\beta}(z_0, p) u^\beta > 0$  ( $< 0$ ) for all pairs  $u, p$  of orthogonal unit vectors.*

(6.2) *The function  $F(z, z')$  is positive (negative) quasi-regular at  $z_0$  if  $u^a F_{\alpha\beta}(z_0, p) u^\beta \geq 0$  ( $\leq 0$ ) for all pairs  $u, p$  of orthogonal unit vectors.*

We use the abbreviations p.q.r., n.q.r. for positive quasi-regular, negative quasi-regular, respectively. It is well known that  $F(z, z')$  is p.q.r. (n.q.r.) at  $z_0$  if and only if  $\mathcal{E}(z_0, p, r) \geq 0$  ( $\leq 0$ ) for all  $p \neq 0$  and all  $r$ . Although we do not use the concept in this note, we shall also make the following definition:

(6.3) *The function  $F(z, z')$  is p.q.r. normal (n.q.r. normal) at  $z_0$  if  $\mathcal{E}(z_0, p, r) > 0$  ( $< 0$ ) whenever  $p \neq 0$  and  $r \neq kp$ ,  $k \geq 0$ .*

With this terminology we state the following lemma:

**LEMMA 5.** *If  $\phi'_0(t_0)$  and  $\phi'(t_0)$  are defined, and  $F(z, z')$  is p.q.r. at  $z_0(t_0)$ , then  $\phi'(t_0) \geq \phi'_0(t_0)$ .*

Let  $\epsilon$  be an arbitrary positive number, and let  $F^\epsilon(z, z') = F(z, z') + \epsilon|z'|$ . By an elementary computation, we find that if  $u$  and  $p$  are orthogonal unit vectors, then

$$(6.4) \quad u^a F_{\alpha\beta}^\epsilon(z, p) u^\beta = u^a F_{\alpha\beta}(z, p) u^\beta + \epsilon.$$

The first term on the right is nonnegative at  $z = z_0(t_0)$ , by hypothesis. Hence the left-hand member is positive for all  $z$  of the set  $S$  in a neighborhood  $U$

†  $F_{ij}(z, z')$  means  $F_{\alpha\beta\gamma\delta}(z, z')$ .

of  $z_0(t_0)$ . If  $h$  is a sufficiently small positive number, less than  $\dagger 1 - t_0$ , then for all large  $n$  the arcs  $z = z_n(t)$  and  $z = z_0(t)$ , ( $t_0 \leq t \leq t_0 + h$ ), are in  $U$ , and by a known semicontinuity theorem

$$(6.5) \quad \liminf \int_{t_0}^{t_0+h} F^*(z_n, \dot{z}_n) dt \geq \int_{t_0}^{t_0+h} F^*(z_0, \dot{z}_0) dt.$$

Recalling the definition of  $F^*$  and the inequality  $|\dot{z}_n| \leq L$ , we deduce that

$$(6.6) \quad \liminf [\phi_n(t_0 + h) - \phi_n(t_0)] + \epsilon h L \geq \phi_0(t_0 + h) - \phi_0(t_0).$$

That is,

$$(6.7) \quad \phi(t_0 + h) - \phi(t_0) + \epsilon h L \geq \phi_0(t_0 + h) - \phi_0(t_0).$$

If we divide by  $h$  and let  $h$  tend to 0, we obtain

$$(6.8) \quad \phi'(t_0) + \epsilon L \geq \phi'_0(t_0).$$

But  $\epsilon$  is an arbitrary positive number; so (6.8) implies

$$(6.9) \quad \phi'(t_0) \geq \phi'_0(t_0),$$

which was to be proved.

Our next theorem is the following:

**THEOREM 2.** *Let  $S$  be a closed point set in  $q$ -dimensional space. If  $F(z, z')$  satisfies condition (2.1), and condition (4.1) holds at every interior point of  $S$ , while  $F(z, z')$  is positive quasi-regular at each boundary point of  $S$ , then in the class  $K$  there is a curve which minimizes  $\mathcal{J}(C)$ .*

Choose first a minimizing sequence  $C_n^*$  of curves  $z = z_n^*(t)$ , ( $0 \leq t \leq 1$ ). There is a set of values of  $t$  open relative to  $[0, 1]$  for which  $z_n^*(t)$  has distance greater than  $1/n$  from the boundary of  $S$ . This open set consists of a finite or denumerable number of intervals, open relative to  $[0, 1]$ . Only a finite number  $h_n$  of these define arcs  $C_{n,j}^*$ , ( $j = 1, \dots, h_n$ ), of  $C_n^*$  of length greater than  $1/n$ . These arcs we replace by polygonal arcs  $\Pi_{n,j}^*$  having the same end points and such that  $|\mathcal{J}(\Pi_{n,j}^*) - \mathcal{J}(C_{n,j}^*)| < 1/nh_n$ , ( $j = 1, \dots, h_n$ ). If we denote by  $C_n'$  the curve obtained by replacing the arcs  $C_{n,j}^*$  by the polygonal arcs  $\Pi_{n,j}^*$ , then  $|\mathcal{J}(C_n') - \mathcal{J}(C_n^*)| < 1/n$ ; so  $\lim \mathcal{J}(C_n') = \mu$ . By a minor modification of Lemma 1, each arc  $\Pi_{n,j}^*$  can be replaced by an arc  $\Pi_{n,j}$  having a number  $s_{n,j}$  of vertices which is not greater than the number of vertices of  $\Pi_{n,j}^*$  and which minimizes  $\mathcal{J}(C)$  in the class of polygons joining the ends of  $\Pi_{n,j}^*$  and having not more than  $s_{n,j}$  vertices. Let  $C_n$  be the curve obtained by replacing the arcs  $\Pi_{n,j}^*$  by the arcs  $\Pi_{n,j}$ . Then  $\lim \mathcal{J}(C_n) = \mu$ .

$\dagger$  If  $t_0 = 1$ , we take  $h$  negative; modifications are obvious.

Now let  $t_0$  be a point of  $E$ . If  $z_0(t_0)$  is interior to  $S$ , there is a neighborhood of  $z_0(t_0)$  on which the  $C_n$  are polygonal if  $n$  is large. All the discussion leading up to equation (4.7) remains valid, for Lemmas 3 and 4 are purely local. So in this case (6.9) is valid. If  $z_0(t_0)$  is a boundary point of  $S$ , then by Lemma 5 inequality (6.9) holds. Integrating, we obtain

$$(6.10) \quad \mu = \phi(1) \geq \phi_0(1) = \mathcal{J}(C_0).$$

But  $C_0$  is in  $K$ ; so  $\mathcal{J}(C_0) \geq \mu$ . This, with (6.10), proves  $\mathcal{J}(C_0) = \mu$ , and the theorem is established.

**7. Geometric interpretation of approach sets.** For fixed  $z$ , let us construct the graph in  $(u, p)$ -space of the function  $u = F(z, p)$ . As is well known, this is a conical hypersurface with vertex at the origin. Let  $A$  be an approach set at  $z$ . If  $p_1$  and  $p_2$  are both in  $A$ , then

$$(7.1) \quad F_i(z, p_1) = F_i(z, p_2), \quad i = 1, \dots, q.$$

But  $u = p^\alpha F_\alpha(z, p_i)$  is the equation of the hyperplane tangent to  $u = F(z, p)$  at  $p_i$ , ( $j = 1, 2$ ). Therefore equation (7.1) shows that this same hyperplane is tangent to the surface  $u = F(z, p)$  at all points of  $A$ .

Conversely, let  $u = l_\alpha p^\alpha$  be a hyperplane tangent to the hypersurface  $u = F(z, p)$  for all  $p$  in a set  $A$ . Then if  $p$  is in  $A$ , the partial derivatives  $F_i(z, p)$  must be the same as the partial derivatives  $l_i$  of the osculating function  $l_\alpha p^\alpha$ . Therefore  $F_i(z, p) = l_i$  for all  $p$  in  $A$ , and  $A$  is an approach set at  $z$ .

If  $F(z, p) > 0$  for  $|p| > 0$ , this interpretation can be formulated somewhat differently. In the preceding paragraphs set  $u = 1$ . The hyperplane  $u = 1$  intersects  $u = F(z, p)$  in a  $(q-1)$ -dimensional hypersurface, and the hyperplane  $u = l_\alpha p^\alpha$  is tangent to  $F(z, p) = 1$  if and only if  $1 = l_\alpha p^\alpha$  is tangent to  $F(z, p) = 1$ . Thus if  $q = 2$  and  $F(z, p) > 0$  for  $|p| > 0$ , we can find all approach sets at  $z$  by constructing the curve  $F(z, p) = 1$  and finding all the points  $(p^1, p^2)$  at which an arbitrary line  $a_1 p^1 + a_2 p^2 + a_3 = 0$  is tangent to the curve.

This interpretation suggests that it might be more descriptive to replace the name "approach set" by "isotangential set."

**8.  $\mathcal{E}$ -admissibility of approach sets; a more general existence theorem.** The geometric interpretation in §7 suggests the following line of reasoning. If  $z = z(t)$  minimizes  $\mathcal{J}(C)$ , then for each  $t_0$  the surface  $u = F(z(t_0), p)$  never sinks below the plane tangent to it at  $p = z'(t_0)$ . The analytical statement is  $\mathcal{E}(z(t_0), z'(t_0), p) \geq 0$  for all  $p$ . The definition of approach sets  $A$  was suggested by the Weierstrass-Erdman corner condition. Might it not be true that only those approach sets  $A$  are of importance which have the property that the hypersurface  $u = F(z, p)$  never sinks below the hyperplane tangent to it at

the points  $p$  of  $A$ ? We shall give this property of the set  $A$  the name " $\mathcal{E}$ -admissibility." Stated analytically the definition is as follows:

(8.1) *If  $A$  is an approach set at  $z$ , it is  $\mathcal{E}$ -admissible if  $\mathcal{E}(z, p_0, p) \geq 0$  for all  $p_0$  in  $A$  and all  $p$ .*

It would make no difference here if we replaced the words "all  $p_0$  in  $A$ " by the words "some one  $p_0$  in  $A$ ." For if  $p_0, p_1$  are any two vectors in  $A$ , then for all  $p$  the equation

$$(8.2) \quad \begin{aligned} \mathcal{E}(z, p_0, p) &= F(z, p) - p^\alpha F_\alpha(z, p_0) = F(z, p) - p^\alpha F_\alpha(z, p_1) \\ &= \mathcal{E}(z, p_1, p) \end{aligned}$$

holds.

The conjecture that approach sets which are not  $\mathcal{E}$ -admissible can be disregarded is indeed true, and we shall establish it in Theorem 3 below. However, the proof is rather complicated. It does not seem possible to rest upon the minimizing property of each  $\Pi_n$ , as we have done before. For in any direct analogue of the proof of the Weierstrass condition polygons are introduced which have new vertices, and  $\Pi_n$  does not necessarily minimize  $\mathcal{Y}(C)$  in the class of polygons with not more than  $s_n + 1$  vertices. The proof which we shall give is therefore based on the property of the sequence as a whole that  $\lim \mathcal{Y}(\Pi_n) = \mu$ .

**THEOREM 3.** *Theorems 1 and 2 remain valid if in hypothesis (4.1) the words "each approach set" are replaced by "each  $\mathcal{E}$ -admissible approach set."*

In the proof of Theorems 1 and 2 the approach set  $A$  entered by way of Lemma 4, and then only in case there was no subsequence  $\{\Pi_k\}$  and  $\delta > 0$  for which either (a)  $\lim_{k \rightarrow \infty} \theta(h, t_0 - \delta, t_0 + \delta) > t_0$  or (b)  $\lim_{k \rightarrow \infty} \Theta(h, t_0 - \delta, t_0 + \delta) \leq t_0$ . For if either (a) or (b) held, the proof was relatively simple and did not involve any approach set  $A$ . We may suppose then that there is no subsequence  $\{\Pi_k\}$  and no  $\delta > 0$  for which either (a) or (b) holds. If we can then show that the approach set  $A$  in Lemma 4 is necessarily  $\mathcal{E}$ -admissible, our proof is complete.

In Lemma 4 there is no loss of generality in assuming that the common value of the partial derivatives  $F_\alpha(z_0(t_0), p)$  for all  $p$  in  $A$  is zero. For let us replace  $F(z, p)$  by  $F(z, p) - p^\alpha F_\alpha(z_0(t_0), p_0)$ , where  $p_0$  is in  $A$ . All the curves  $C$  of the family  $K$  join  $z_1$  to  $z_2$ ; so  $\mathcal{Y}(C)$  changes by  $(z_2^\alpha - z_1^\alpha) F_\alpha(z_0(t_0), p_0)$ , independent of  $C$ . Therefore the minimizing properties of the  $\Pi_n$  are unchanged by the alteration of  $F(z, p)$ . The statement that  $A$  is not  $\mathcal{E}$ -admissible assumes the form that there is a  $p_1$  such that  $F(z_0(t_0), p_1) < 0$ . Because of the homogeneity of  $F$ , we may suppose that  $p_1$  is a unit vector, and we write

$$F(z_0(t_0), p_1) = -\gamma < 0.$$

By the continuity of  $F$ , there is an  $\epsilon > 0$  such that

$$(8.3) \quad F(z, p_1) < -6\gamma$$

if  $|z - z_0(t_0)| < 4\epsilon$ . Furthermore,  $F_i(z_0(t_0), p) = 0$  if  $p$  is in  $A$ ; hence we may suppose that  $\epsilon$  has been chosen small enough so that the following statement holds:

(8.4) *The vector  $(F_1(z, p), \dots, F_q(z, p))$  has length less than  $\gamma$  if  $|z - z_0(t_0)| < 4\epsilon$ ,  $|z_2 - z_1| \leq |p| \leq 2L + 1$ , and  $p$  is in the  $3\epsilon$ -neighborhood of  $A$ .*

Clearly we may suppose  $\epsilon < \min [1, |z_2 - z_1|]$ .

On the bounded closed set  $|z - z_0(t_0)| \leq 1$ ,  $|p| \leq 2L + 1$ , the functions  $F_{i^*}(z, p)$  are continuous; hence we have the following statement:

(8.5) *There is an  $M > 0$  such that the vector  $(F_{i^*}(z, p), \dots, F_{s^*}(z, p))$  has length less than  $M$  if  $|z - z_0(t_0)| \leq 1$  and  $|p| \leq 2L + 1$ .*

Now from the subsequence  $\{\Pi_m\}$  of Lemma 4 we choose a subsequence  $\{\Pi_h\}$  with the following properties:

(8.6)  $|z_h(t) - z_0(t)| < \epsilon$  for  $0 \leq t \leq 1$  and all  $h$ .

(8.7) The limits  $t_1 = \lim_{h \rightarrow \infty} \theta(h, t_0 - \delta, t_0 + \delta)$  and  $t_2 = \lim_{h \rightarrow \infty} \Theta(h, t_0 - \delta, t_0 + \delta)$  exist, where  $\delta$  is prescribed by Lemma 4.

(As previously remarked, we need consider only the case  $t_1 \leq t_0 < t_2$ .)

(8.8)  $\delta$  is chosen less than  $\epsilon/L$ .

(8.9) For all  $h$  the following inequality holds:

$$\Theta(h, t_0 - \delta, t_0 + \delta) - \theta(h, t_0 - \delta, t_0 + \delta) > (t_2 - t_1)/2.$$

Each of the conditions (8.6), (8.7), (8.9) is readily satisfied by appropriate choice of a subsequence; for (8.6) and (8.9) we need only reject a finite number of terms.

Let  $d_{h,1}, \dots, d_{h,n(h)}$  be a sequence of points with the following properties:

(8.10)  $d_{h,1} < d_{h,2} < \dots$

(8.11)  $|d_{h,j+1} - d_{h,j}| < \gamma/M$  for  $j = 1, \dots, n(h) - 1$ .

(8.12)  $d_{h,1} = \theta(h, t_0 - \delta, t_0 + \delta)$ ,  $d_{h,n(h)} = \Theta(h, t_0 - \delta, t_0 + \delta)$ .

(8.13) Every  $t$  between  $\theta(h, t_0 - \delta, t_0 + \delta)$  and  $\Theta(h, t_0 - \delta, t_0 + \delta)$  which defines a vertex of  $\Pi_h$  is included among the  $d_{h,j}$ . (Thus  $z_h(t)$  is linear on  $d_{h,j} \leq t \leq d_{h,j+1}$ .)

Let  $m_{h,j}$  be the mid-point of the interval  $[d_{h,j}, d_{h,j+1}]$ .

On each interval  $[d_{h,j}, d_{h,j+1}]$  we shall replace the line segment  $z = z_h(t)$  belonging to  $\Pi_h$  by a polygon of two sides having the same beginning and the same end as the line segment. To keep the notation from becoming too complicated, we temporarily replace the symbols  $d_{h,j}$ ,  $m_{h,j}$ ,  $d_{h,j+1}$ ,  $z_h$  by  $d_1$ ,  $m$ ,  $d_2$ ,  $z$ ,

respectively. For  $0 \leq \tau \leq \epsilon$  we define  $\Pi(\tau)$  to be the two-sided polygon  $z = z(t, \tau)$ , where

$$(8.14) \quad \begin{aligned} z(t, \tau) &= z(d_1) + (t - d_1)\tau p_1, & d_1 \leq t < m, \\ z(t, \tau) &= z(2t - d_2) + (d_2 - t)\tau p_1, & m \leq t \leq d_2. \end{aligned}$$

We readily verify that  $z(t, \tau)$  is continuous on  $[d_1, d_2]$  and that

$$(8.15) \quad z(d_1, \tau) \equiv z(d_1), \quad z(d_2, \tau) \equiv z(d_2).$$

Also it is easy to show that

$$(8.16) \quad \int_m^d F(z(t, 0), z'(t, 0)) dt = \int_{d_1}^{d_2} F(z(t), z'(t)) dt,$$

where the prime denotes the derivative with respect to  $t$ .

The polygon  $\Pi(\tau)$  may be regarded as arising from  $\Pi(0)$  by displacing the vertex  $z(m, 0)$  by the amount  $(d_2 - d_1)\tau p_1/2$ . Hence we may calculate  $\mathcal{Y}'(\Pi(\tau))$  by applying (2.8) of I to the sides  $[d_1, m]$ ,  $[m, d_2]$  and adding. We obtain

$$(8.17) \quad \begin{aligned} \frac{d}{d\tau} \mathcal{Y}(\Pi(\tau)) &= \frac{1}{2}(d_2 - d_1)F_a(z(\bar{t}, \tau), p_1)p_1^\alpha \\ &\quad - \frac{1}{2}(d_2 - d_1)F_a(z(\bar{t}, \tau), z'(\bar{t}, \tau))p_1^\alpha \\ &\quad + \int_{d_1}^{d_2} F_{z''}(z(t, \tau), z'(t, \tau))p_1^\alpha |m - t| dt, \end{aligned}$$

where  $d_1 < \bar{t} < m$  and  $m < \bar{t} < d_2$ . Now we must prove the following statements:

$$(8.18) \quad \text{If } d_1 \leq t \leq d_2 \text{ and } 0 \leq \tau \leq \epsilon, \text{ then } |z(t, \tau) - z_0(t_0)| < 3\epsilon.$$

$$(8.19) \quad \text{If } m < t < d_2 \text{ and } 0 \leq \tau \leq \epsilon, \text{ then } z'(t, \tau) \text{ is in the } \epsilon\text{-neighborhood of } A, \\ \text{and } |z_2 - z_1| \leq |z'(t, \tau)| \leq 2L + 1.$$

We establish (8.19) first. By (8.14), we have

$$z'(t, \tau) = 2z'(2t - d_2) - \tau p_1.$$

But by our choice of parameter  $t$ , we have  $|z_2 - z_1| \leq z'(2t - d_2) \leq L$ , while  $|\tau p_1| \leq \epsilon < \min [|z_2 - z_1|, 1]$ . Hence  $|z_2 - z_1| < |z'(t, \tau)| < 2L + 1$ . Also, by Lemma 4 the vector  $z'(2t - d_2)$  is in the  $\epsilon$ -neighborhood of  $A$ ; whence  $z'(t, \tau)$  is in the  $3\epsilon$ -neighborhood of  $A$ . Thus (8.19) is established.

By (c) of Lemma 2,  $|z'(t)| \leq L$ . Hence if  $t_0 - \delta \leq t \leq t_0 + \delta$ , then (recalling (8.8)) we have

$$|z(t) - z(t_0)| \leq L\delta < \epsilon.$$

By (8.6),  $|z(t_0) - z_0(t_0)| < \epsilon$ . Combining these inequalities, we find that both

$z(d_1)$  and  $z(d_2)$  are in the  $2\epsilon$ -neighborhood of  $z_0(t_0)$ . The distance from  $z(d_1) = z(d_1, \tau)$  to  $z(m, \tau)$  is  $|(m - d_1)\tau p_1| < \epsilon$ . So all three vertices  $z(d_1, \tau)$ ,  $z(m, \tau)$ , and  $z(d_2, \tau)$  are in the  $3\epsilon$ -neighborhood of  $z_0(t_0)$ , and (8.18) is proved.

Thus the arguments in (8.17) satisfy the requirements laid down in (8.3), (8.4), and (8.5), and we have from (8.17) and (8.11)

$$(8.20) \quad \begin{aligned} \frac{d}{d\tau} \mathcal{F}(\Pi(\tau)) &\leq -3\gamma(d_2 - d_1) + \frac{1}{2}\gamma(d_2 - d_1) + \int_{d_1}^{d_2} M |m - t| dt \\ &< -2\gamma(d_2 - d_1) + M(d_2 - d_1)^2 < -\gamma(d_2 - d_1). \end{aligned}$$

If we integrate from  $\tau = 0$  to  $\tau = \epsilon$ , we obtain (using (8.16))

$$(8.21) \quad \mathcal{F}(\Pi(\epsilon)) < \mathcal{F}(\Pi(0)) - \gamma\epsilon(d_2 - d_1) = \mathcal{F}(\Pi) - \gamma\epsilon(d_2 - d_1).$$

In the foregoing paragraphs  $[d_1, d_2]$  was any one of the intervals  $[d_{h,i}, d_{h,i+1}]$ . Let the construction be carried out on each of these intervals. The arc  $z = z_h(t)$ ,  $d_{h,1} \leq t \leq d_{h,n(h)}$  is thereby replaced by a polygonal arc  $z = z_h(t, \epsilon)$  having the same ends and satisfying the relation

$$\int_{d_{h,1}}^{d_{h,n(h)}} F(z_h(t, \epsilon), \dot{z}_h(t, \epsilon)) dt < \int_{d_{h,1}}^{d_{h,n(h)}} F(z_h, \dot{z}_h) dt - \sum_{j=1}^{n(h)-1} \gamma\epsilon(d_{h,i+1} - d_{h,i}).$$

If we extend the range of definition of  $z_h(t, \epsilon)$  by setting it equal to  $z_h(t)$  for  $0 \leq t < d_{h,1}$  and for  $d_{h,n(h)} < t \leq 1$ , we obtain a polygon  $\Pi_h(\epsilon)$  joining  $z_1$  to  $z_2$  and satisfying (by (8.21) and (8.9))

$$\begin{aligned} \mathcal{F}(\Pi_h(\epsilon)) &< \mathcal{F}(\Pi_h) - \sum_{j=1}^{n(h)-1} \gamma\epsilon(d_{h,i+1} - d_{h,i}) = \mathcal{F}(\Pi_h) - \gamma\epsilon(d_{h,n(h)} - d_{h,1}) \\ &< \mathcal{F}(\Pi_h) - \gamma\epsilon(t_2 - t_1)/2. \end{aligned}$$

Therefore

$$\limsup_{h \rightarrow \infty} \mathcal{F}(\Pi_h(\epsilon)) \leq \lim_{h \rightarrow \infty} \mathcal{F}(\Pi_h) - \gamma\epsilon(t_2 - t_1)/2 = \mu - \frac{1}{2}\gamma\epsilon(t_2 - t_1) < \mu.$$

This is impossible by the definition of  $\mu$ , and our theorem is established.

It is evident that the restriction  $z_2 \neq z_1$  can be removed here just as it was in §6. We thus arrive at the following theorem, which includes the three preceding ones as special cases:

**THEOREM 4.** Let  $F(z, z')$  be defined and satisfy the usual continuity and homogeneity conditions\* for all  $z$  in a closed set  $S$  and all  $z'$ . Let  $F(z, z')$  satisfy the following conditions:

- (i) At every boundary point  $z$  of  $S$ ,  $F(z, z')$  is positive quasi-regular.
- (ii) At every interior point  $z$  of  $S$ , every  $\mathcal{E}$ -admissible approach set  $A$  is

\* As specified in I.

the sum of a finite number of convex sets  $A_1, \dots, A_k$  such that  $\Omega(z, p_i, p_i) < 0$  if  $p_i$  is in  $A_i$ ,  $p_i$  in  $A_j$ ,  $i < j$ .

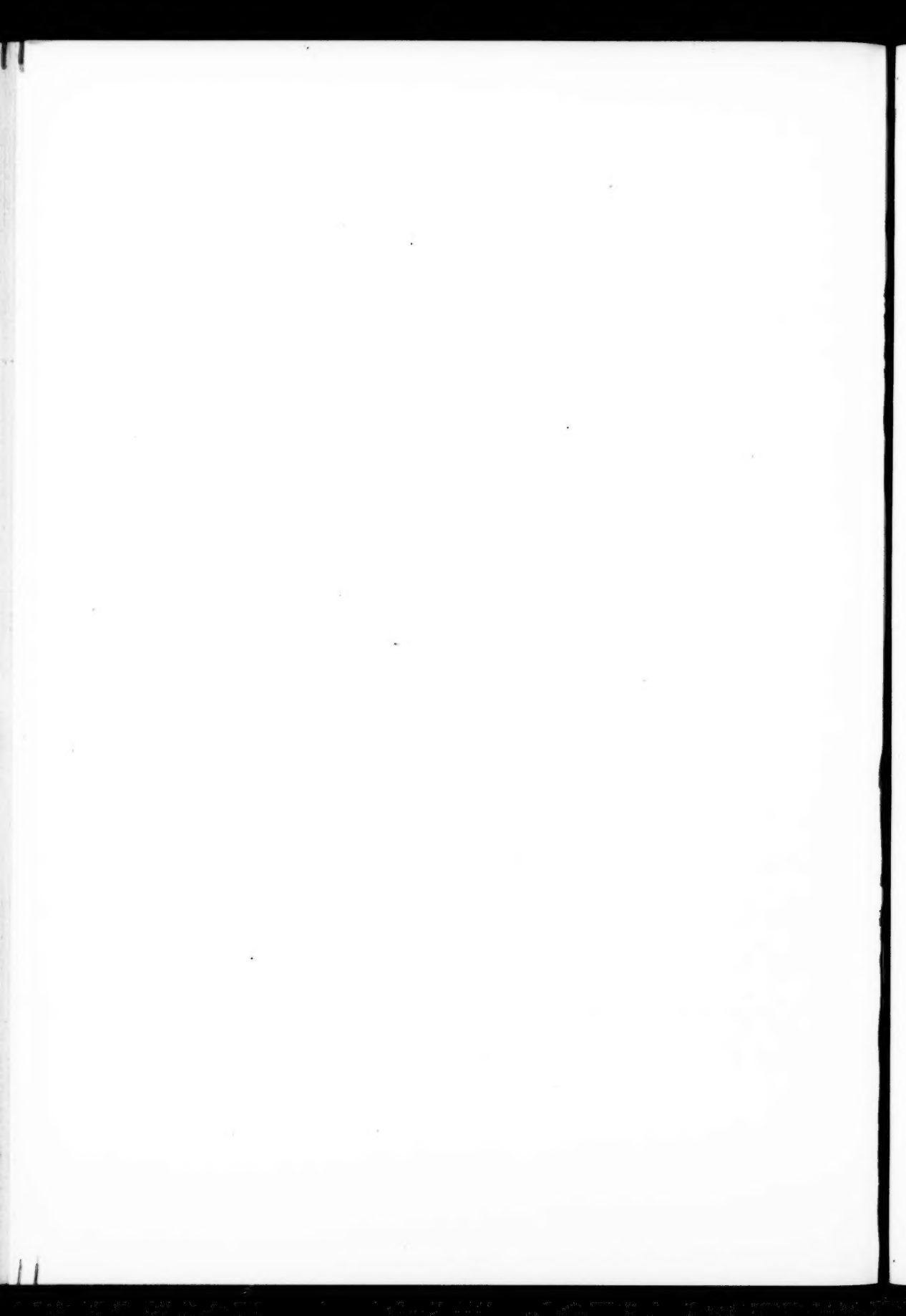
Let condition (2.1) be satisfied. Then for every two points  $z_1, z_2$  of  $S$  the class  $K$  of all rectifiable curves lying in  $S$  and joining  $z_2$  to  $z_1$  either is empty or contains a minimizing curve for  $\mathcal{F}(C)$ .

**9. Geometric interpretation of  $\mathcal{E}$ -admissibility.** The geometric interpretation of the property of  $\mathcal{E}$ -admissibility has already been given at the beginning of §8. If  $A$  is an approach set at  $z$ , it is  $\mathcal{E}$ -admissible if and only if the surface  $u = F(z, p)$  never sinks below the hyperplane tangent to it at the points  $(u, p)$  with  $p$  in  $A$ . The alternative geometric interpretation of approach sets also carries with it an interpretation of  $\mathcal{E}$ -admissibility. Let the hyperplane (in  $p$ -space)  $a_p p^a + b = 0$  be tangent to the hypersurface  $F(z, p) = 1$  at all the points  $p$  of the surface which belong to  $A$ . This hyperplane divides  $p$ -space into two half-spaces. The half-space which contains the origin necessarily contains other points of  $F(z, p) = 1$ . The set  $A$  is then  $\mathcal{E}$ -admissible if and only if the entire hypersurface  $F = 1$  lies in this same half-space.

It follows that if we form the least convex body  $Q$  containing  $F(z, p) = 1$ , then the points  $p$  of the hypersurface which are not on the boundary of  $Q$  cannot belong to any  $\mathcal{E}$ -admissible approach set. In particular, if  $q = 2$ , the points  $p_0$  of  $F(z, p) = 1$  on the boundary of  $Q$  either lie in a line segment belonging to the boundary of  $Q$  or they do not. In the second case the approach set  $A$  containing  $p_0$  consists of the half-line from the origin through  $p_0$ . In the first case, let  $B$  be the set of all points  $p$  at which the line segment touches  $F(z, p) = 1$ . The (maximal) approach set at  $z$  containing  $p_0$  then consists of all the half-lines from the origin through the points of  $B$ .

Consider, for example, the function  $F(x, y, x', y')$  studied in §5. If  $|y| \leq 1/2$ , the curve  $F = 1$  is convex. If  $|y| > 1/2$ , the curve is dumbbell-shaped, curved inwards near its intersections with the  $x'$ -axis. It is then evident from the graph that there are just two approach sets which consist of more than one half-line, and these are determined by the points of contact with the two lines  $x' = \text{const.}$  tangent to the curve at the points at which  $x'$  is respectively greatest and least. In §5 we showed rather more than this. We showed that these are the *only* approach sets other than half-lines; that besides these there are no other approach sets other than half-lines, whether  $\mathcal{E}$ -admissible or not. But Theorem 3 shows that the simpler conclusion reached here is sufficient to show the existence of minimizing curves for  $\mathcal{F}(C)$ .

UNIVERSITY OF VIRGINIA,  
CHARLOTTESVILLE, VA.



## SOME EXISTENCE THEOREMS IN THE CALCULUS OF VARIATIONS

### IV. ISOPERIMETRIC PROBLEMS IN NON-PARAMETRIC FORM\*

BY

E. J. McSHANE

1. **Heuristic outline of proof.** The preceding notes in this series have already appeared in these Transactions,<sup>†</sup> and will be cited as I, II, III, respectively.

Here we shall consider isoperimetric problems with one side integral, in the non-parametric form. A heuristic outline of the proof is as follows. We seek to minimize the integral

$$(1.1) \quad \mathcal{J}[y] = \int_{x_0}^X f(x, y, \dot{y}) dx$$

in the class  $K$  of a.c. curves  $y=y(x)$  joining two fixed points  $(x_0, y_0)$  and  $(X, Y)$  and giving a fixed value  $\gamma$  to the integral

$$(1.2) \quad \mathcal{G}[y] = \int_{x_0}^X g(x, y, \dot{y}) dx.$$

Under suitable conditions, we can use as minimizing sequence a sequence of polygons  $\Pi_n: y=y_n(x)$  such that  $\mathcal{J}[y_n]$  tends to its lower bound  $\mu$  on the class  $K$  and  $\gamma_n = \mathcal{G}[y_n]$  tends to  $\gamma$ . Also we may assume that if  $s_n$  is the number of vertices of  $\Pi_n$ , then  $\Pi_n$  is a "best" polygon, in the sense that in the class of polygons of not more than  $s_n$  vertices joining  $(x_0, y_0)$  to  $(X, Y)$  and giving the value  $\gamma_n$  to  $\mathcal{G}[y]$ , the polygon  $\Pi_n$  minimizes  $\mathcal{J}[y]$ .

There are known conditions which must be satisfied by a curve  $\bar{C}: y=\bar{y}(x)$ ,  $(x_0 \leq x \leq X)$ , which is of class  $D'$  and minimizes  $\mathcal{J}(C)$  on the class  $K$ . For our purposes the statement of these conditions is more conveniently written by passing to the parametric notation, by equation (1.1) of I. That is, we write

---

\* Presented to the Society, April 15, 1938; received by the editors October 29, 1937 and, in revised form, February 16, 1938.

† I. *The Dresden corner condition*, these Transactions, vol. 44 (1938), pp. 429-438; II. *Existence theorems for isoperimetric problems in the plane*, *ibid.*, pp. 439-453; III. *Existence theorems for nonregular problems*, these Transactions, vol. 45 (1939), pp. 151-171.

$$(1.3) \quad \begin{aligned} F(z, z') &= F(z^0, z^1, \dots, z^q, z^{0'}, \dots, z^{q'}) \\ &= z^{0'} f(z^0, z^1, \dots, z^q, z^{1'}/z^{0'}, \dots, z^{q'}/z^{0'}) \end{aligned}$$

for  $z^{0'} > 0$ . Correspondingly we write  $\bar{C}$  in the form  $z = \bar{z}(t)$ ,  $(0 \leq t \leq 1)$ . Then it is known that there are constants  $\lambda$  and  $\lambda_0$ , not both zero, such that the DuBois-Reymond relation holds:

$$(1.4) \quad H_i(\bar{z}(t), \bar{z}'(t)) = \int_0^t H_{,i}(\bar{z}(t), \bar{z}'(t)) dt + c_i, \quad i = 0, 1, \dots, q,$$

where  $H = \lambda_0 F - \lambda G$ . If we suppose that  $\bar{C}$  is a normal curve, so that  $\lambda_0 \neq 0$ , we can suppose  $\lambda_0 = 1$ . Let  $\bar{i}$  define a corner of  $\bar{C}$ , and write  $\bar{p} = \bar{z}'(\bar{i}-0)$ ,  $\bar{r} = \bar{z}'(\bar{i}+0)$ . By the Dresden corner condition\*

$$(1.5) \quad \Omega_H(\bar{z}(\bar{i}), \bar{p}, \bar{r}) \leq 0.$$

Also, from (1.4) we derive the Weierstrass-Erdman corner condition:

$$(1.6) \quad H_i(\bar{z}(\bar{i}), \bar{p}) = H_i(\bar{z}(\bar{i}), \bar{r}), \quad i = 0, 1, \dots, q.$$

Of course we cannot expect the polygons  $\Pi_n$  to satisfy these conditions. Nevertheless, each  $\Pi_n$  does minimize  $\mathcal{Y}(C)$  in a certain class of polygons with  $G(C) = \text{const.}$ , and it is therefore not unreasonable to hope that some approximate form of (1.4), (1.5), and (1.6) is satisfied by each  $\Pi_n$ .

Condition (1.6) can be written in a way which does not involve  $\lambda_0$  and  $\lambda$ . We make the following definition:

(1.7) *A set  $A$  of (nonzero) vectors is an approach set at  $z$  if the matrix*

$$\Delta(z, p, r) \equiv \begin{pmatrix} F_i(z, p) - F_i(z, r) \\ G_i(z, p) - G_i(z, r) \end{pmatrix}, \quad i = 0, 1, \dots, q,$$

*has rank less than 2 for each pair  $p, r$  of vectors belonging to  $A$ .*

Then  $\bar{p}$  and  $\bar{r}$  belong to an approach set at  $\bar{z}(\bar{i})$ . For by (1.6) a linear combination of the rows ( $\lambda_0$  times the first row minus  $\lambda$  times the second) vanishes. We shall shortly see that for a fairly large class of functions  $G$  each approach set  $A$  at  $z$  determines a pair of numbers  $\lambda_0 = 1$  and  $\lambda$  for which (1.6) holds.

We now proceed to over-simplify the situation by proceeding as though conditions (1.4), (1.5), and (1.6) actually hold on each  $\Pi_n$ . Furthermore, we assume that at each point  $z$  each approach set contains at most two unit vectors  $p, r$ , and  $\Omega_H(z, p, r) \neq 0$ . We suppose that the sequence  $z_n(t)$  of functions defining the polygons  $\Pi_n$  converges uniformly to a limit function  $z_0(t)$ ;

\* I, Theorem 2.

this is not hard to bring about. Consider a particular value  $t_0$  of  $t$  and a subsequence  $\{\Pi_m\}$  for which  $z'_m(t_0)/|z'_m(t_0)|$  tends to a limit  $p$ . If  $\{t_m\}$  is any sequence of points approaching  $t$ , then by (1.4) the difference

$$H_i(z_m(t_m), z'_m(t_m)) - H_i(z_m(t_0), z'_m(t_0))$$

tends to zero as  $m \rightarrow \infty$ , since it is the integral of a bounded function over an interval which is shrinking to zero. Hence as  $m \rightarrow \infty$  the vector  $z'_m(t_m)/|z'_m(t_m)|$ , if it converges to any limit, converges to a limit  $r$  for which  $H_i(z_0(t_0), r) = H_i(z_0(t_0), p)$ , ( $i=0, 1, \dots, q$ ); that is,  $p$  and  $r$  belong to an approach set at  $z_0(t_0)$ . In other words, if we choose a small arc of  $\Pi_m$  corresponding to  $t$  near  $t_0$ , each side of  $\Pi_m$  has direction either near  $p$  or near  $r$ . To be specific, we assume  $\Omega_H(z_0(t_0), p, r) < 0$ . Then no side of  $\Pi_m$  with direction near  $r$  can be followed by one with direction near  $p$ . For if this happened, at such a corner  $\bar{i}$  we would have  $\Omega_H(z_m(\bar{i}), z'_m(\bar{i}-0), z'_m(\bar{i}+0))$  near  $\Omega_H(z_0(t_0), r, p)$ , which is positive; and this contradicts (1.5). Therefore the arc of  $\Pi_m$  near  $z_0(t_0)$  consists of a succession of sides with directions near  $p$  followed by a succession of sides with directions near  $r$ . That is, the arc consists of two subarcs, each almost linear. This makes it highly plausible that (as would certainly be the case if both subarcs were actually linear) the integrals of  $F$  and  $G$  along this arc of  $\Pi_m$  tend, respectively, to the integrals of  $F$  and  $G$  along the arc of  $z=z_0(t)$  to which they tend. If this could be applied to a succession of subarcs covering the whole range of  $t$ , it would follow that  $\mathcal{Y}(\Pi_n)$  and  $\mathcal{G}(\Pi_n)$  tended, respectively, to  $\mathcal{Y}(C_0)$  and  $\mathcal{G}(C_0)$ , and  $C_0$  would be the solution sought.

This outline is, of course, over-simplified, and in order to make the proof rigorous it is necessary to depart from it somewhat and add a mass of analytical details. Nevertheless the generalization (8.1) of the assumption that  $\Omega_H(z, p, r) \neq 0$  if  $p, r$  satisfy (1.7) is the vital essence of the proof.

Although we are here concerned with integrals in non-parametric form, in the next paper we shall study integrals in parametric form by closely similar methods. Therefore the lemmas of this paper will be so arranged as to apply simultaneously to both forms of the problem. The chief differences are caused by the requirement  $z^{0'} > 0$  which occurs in studying non-parametric problems. Consequently I shall use the device of enclosing certain statements in square brackets; these are needed if the lemma is understood as applying to non-parametric problems, but should be disregarded for problems in parametric form.

**2. A property of approach sets.** Let us suppose that we are given two integrands  $F(z, z')$ ,  $G(z, z')$  and two distinct fixed points  $z_1, z_2$  in the space of points  $(z^0, z^1, \dots, z^q)$ . We make the following definition:

(2.1) *K is the class of all rectifiable curves C joining  $z_1$  to  $z_2$ , and  $K[G=l]$  is the subclass of K consisting of those curves C of K for which  $G(C)=l$ .*

[In case  $F$  and  $G$  arise from a non-parametric problem, we must restrict our attention to curves representable in the form  $z^i = z^i(z^0)$ , ( $i=1, \dots, q$ ). In this case we define  $K$  as follows:

(2.2) *K is the class of all curves having a.c. representations  $z^i = z^i(z^0)$ , ( $i=1, \dots, q$ ), and joining two fixed points  $z_1, z_2$  with  $z_1^0 < z_2^0$ .  $K[G=l]$  is the subclass of K consisting of those curves C for which  $G(C)=l$ .*

The points  $z_1, z_2$  will also be denoted by  $(x_0, y_0)$ ,  $(X, Y)$ , respectively.]

The problem considered is that of minimizing  $\mathcal{F}(C)$  on the class  $K[G=l]$ .

[We shall now make a hypothesis in the nature of a continuity requirement on  $F$  and  $G$ .

(2.3) *If, as  $n \rightarrow \infty$ , the points  $z_{i,n}, \bar{z}_{i,n}, \zeta_{i,n}$ , and  $\bar{\zeta}_{i,n}$  all converge to a common limit  $z$ , ( $i=0, 1, \dots, q$ ), the numbers  $\theta_{i,n}$  and  $\bar{\theta}_{i,n}$  tend to 0, ( $i=0, 1, \dots, q$ ),  $p_n$  tends to a vector  $p$  with  $p^0 > 0$ , and  $r_n$  tends to a vector  $r$  with  $|r| > 0$ , and if for each  $n$  the matrix*

$$(2.4) \quad \begin{pmatrix} F_i(\bar{z}_{i,n}, r_n) - F_i(z_{i,n}, p_n) + \theta_{i,n} \\ G_i(\bar{\zeta}_{i,n}, r_n) - G_i(\zeta_{i,n}, p_n) + \bar{\theta}_{i,n} \end{pmatrix}$$

*has rank less than 2, then  $r^0 > 0$ .*

Then the following lemma is immediate:

LEMMA 1. *If (2.3) holds and the sequences  $z_{i,n}, \bar{z}_{i,n}, \zeta_{i,n}, \bar{\zeta}_{i,n}, \theta_{i,n}, \bar{\theta}_{i,n}, p_n, r_n$  satisfy the conditions of (2.3), then  $\Delta(z, p, r)$  (of (1.7)) has rank less than 2.*

For then all the functions in the matrix (2.4) are continuous at  $(z, p, r)$ , and the elements in the matrix tend, respectively, to the elements of the matrix  $\Delta(z, p, r)$ .

For problems in parametric form the analogue of Lemma 1 is trivial. It is as follows:

LEMMA 1'. *If as  $n \rightarrow \infty$  the points  $z_{i,n}, \bar{z}_{i,n}, \zeta_{i,n}$ , and  $\bar{\zeta}_{i,n}$  all converge to a common limit  $z$ , ( $i=0, 1, \dots, q$ ), the numbers  $\theta_{i,n}$  and  $\bar{\theta}_{i,n}$  tend to 0, ( $i=0, 1, \dots, q$ ), and  $p_n$  and  $r_n$  tend to nonzero limit vectors  $p, r$ , and if for each  $n$  the matrix (2.4) has rank less than two, then the matrix  $\Delta(z, p, r)$  has rank less than two.*

This is an immediate consequence of the continuity of all the elements of the matrix (2.4).

As is customary in isoperimetric problems, we are often interested in linear combinations of the integrands  $F$  and  $G$ . In order to have a notation

for such combinations we adopt the definition: For each number  $\lambda$ ,

$$(2.5) \quad H(z, z'; \lambda) \equiv F(z, z') - \lambda G(z, z').$$

When there is no danger of misunderstanding, we write  $H(z, z')$  in place of  $H(z, z'; \lambda)$ .

Let us make the definition:

(2.6)  $G(z, r)$  is positive quasi-regular normal at  $z_0$  if  $\mathcal{E}_G(z_0, p, r) > 0$  whenever  $|p| > 0$  and  $r \neq kp$ , ( $k \geq 0$ ) [and  $p^0$  and  $r^0$  are positive];  $G(z, r)$  is negative quasi-regular normal at  $z_0$  if  $\mathcal{E}_G(z_0, p, r) < 0$  whenever  $|p| > 0$  and  $r \neq kp$ , ( $k \geq 0$ ) [and  $p^0$  and  $r^0$  are positive];  $G(z, r)$  is quasi-regular normal at  $z_0$  if it is positive or negative quasi-regular normal at  $z_0$ .

Then we state the following interesting useful property of approach sets:

LEMMA 2. If  $G(z, r)$  is quasi-regular normal at  $z$ , and  $A$  is an approach set at  $z$ , there is a number  $\lambda(z, A)$  such that the function  $H(z, r) \equiv H(z, r; \lambda(z, A)) = F(z, r) - \lambda(z, A)G(z, r)$  satisfies the equations

$$(2.7) \quad H_i(z, p) = H_i(z, r), \quad i = 0, 1, \dots, q,$$

for all  $p, r$  in  $A$ .

Since  $H(z, r)$  is positively homogeneous of degree zero in  $r$ , it is enough to establish (2.7) for unit vectors in  $A$ . Moreover, it is enough to consider the case in which  $G(z, r)$  is positive quasi-regular normal at  $z$ ; if it is negative quasi-regular normal at  $z$ , we replace it by  $-G(z, r)$ .

Let  $p_1, p_2, p_3$  be any three distinct unit vectors in  $A$ . The relationship

$$(2.8) \quad a_1[G_i(z, p_3) - G_i(z, p_2)] + a_2[G_i(z, p_1) - G_i(z, p_3)] \\ + a_3[G_i(z, p_2) - G_i(z, p_1)] = 0, \quad i = 0, 1, \dots, q,$$

obviously holds if  $a_1 = a_2 = a_3$ . We shall show that it holds for no other  $a_1, a_2, a_3$ . Suppose (2.8) holds. By a cyclic interchange of the subscripts 1, 2, 3, followed, if necessary, by a change of sign of all three, we can bring about the relationship  $a_1 \leq a_2 \leq a_3$ . If we subtract  $a_2$  from each of the numbers  $a_i$ , (2.8) remains valid and assumes the form

$$(2.9) \quad (a_1 - a_2)[G_i(z, p_3) - G_i(z, p_2)] \\ + (a_3 - a_2)[G_i(z, p_2) - G_i(z, p_1)] = 0, \quad i = 0, \dots, q.$$

Now we multiply by  $p_2^i$  and sum over  $i = 0, \dots, q$ ; this yields

$$(2.10) \quad -(a_1 - a_2)\mathcal{E}_G(z, p_3, p_2) + (a_3 - a_2)\mathcal{E}_G(z, p_1, p_2) = 0.$$

But  $-(a_1 - a_2)$  and  $(a_3 - a_2)$  are nonnegative, and their coefficients are positive; so this is only possible if  $a_1 - a_2 = a_3 - a_2 = 0$ .

Given any two distinct unit vectors  $p, r$  of  $A$ , the matrix  $\Delta(z, p, r)$  is (by definition of  $A$ ) of rank 0 or 1. Since its second row is not 0, its first row is a number  $\lambda(p, r)$  times the second. From the form of  $\Delta$  it is clear that  $\lambda(p, r) = \lambda(r, p)$ . We must show that  $\lambda(p, r)$  has the same value for all  $p, r$  in  $A$ . If  $A$  contains only two distinct unit vectors  $p, r$ , this is at once evident from the equation  $\lambda(p, r) = \lambda(r, p)$ . It remains to establish the formula in case  $A$  contains more than two distinct unit vectors. To do this it is sufficient to show (with above notation) that  $\lambda(p_1, p_3) = \lambda(p_1, p_2)$ . For then if we fix  $p_1$  and  $p_2 \neq p_1$ ,  $\lambda(p_1, p_3)$  has the same value for all unit vectors  $p_3$  in  $A$ . That is,  $\lambda(p_1, r)$  is independent of its second argument. Since  $\lambda(p, r) = \lambda(r, p)$ , it is also independent of its first argument, and has only one value.

Suppose then that  $p_1, p_2$ , and  $p_3$  are distinct unit vectors in  $A$ . By definition of  $\lambda(p, r)$  we have

$$\begin{aligned} F_i(z, p_3) - F_i(z, p_2) &= \lambda(p_3, p_2) [G_i(z, p_3) - G_i(z, p_2)], \\ F_i(z, p_2) - F_i(z, p_1) &= \lambda(p_1, p_2) [G_i(z, p_2) - G_i(z, p_1)], \\ F_i(z, p_1) - F_i(z, p_3) &= \lambda(p_1, p_3) [G_i(z, p_1) - G_i(z, p_3)]. \end{aligned}$$

Adding, we obtain (2.8) with  $a_1 = \lambda(p_3, p_2)$ ,  $a_2 = \lambda(p_1, p_3)$ ,  $a_3 = \lambda(p_1, p_2)$ . Hence by the preceding proof we know that  $\lambda(p_1, p_3) = \lambda(p_1, p_2)$ ; so  $\lambda(p, r)$  has but one value, which we denote by  $\lambda(z, A)$ .

Thus (2.7) holds if  $p, r$  are distinct unit vectors in  $A$ . Obviously it holds if  $p = r$ ; and the proof is complete.

Under the hypotheses of Lemma 2, the set  $A$  is an approach set for  $H(z, r)$  at the point  $z$  as defined in III, for equations (2.7) are exactly the equations (1.5) used there to define an approach set. Therefore we have available a geometric interpretation of approach sets. When  $G(z, r)$  is quasi-regular normal at  $z$  and  $A$  is an approach set at  $z$ , then there is a single hyperplane  $u = l_a r^a$  tangent to the hypersurface  $u = F(z, r) - \lambda(z, A)G(z, r)$  at all points  $(u, r)$  with  $r$  in  $A$ . Conversely, let  $\lambda$  be a number, and let  $A$  be the set of (nonzero) vectors  $r$  such that a given hyperplane  $u = l_a r^a$  is tangent to  $u = H(z, r; \lambda)$  whenever  $r$  is in  $A$ . Then as in §7 of III,  $A$  is an approach set for  $H(z, r; \lambda)$ , and equations (2.7) hold. But equations (2.7) imply that the rank of  $\Delta(z, p, r)$  is less than two; so  $A$  is an approach set at  $z$  as defined in (1.7).

This interpretation shows that any approach set  $A$  containing two or more distinct unit vectors has a unique maximal extension, provided that  $G(z, r)$  is quasi-regular normal at  $z$ . For then any two unit vectors of  $A$  determine  $\lambda(z, A)$ , and hence determine  $H(z, r)$ . Let  $u = l_a r^a$  be the hyperplane tangent to  $u = H(z, r)$  whenever  $r$  is in  $A$ . Let  $A^*$  be the set of all nonzero

vectors  $r$  for which  $u = l_\alpha r^\alpha$  is tangent to  $u = H(z, r)$ . Then  $A^*$  is an approach set and contains  $A$ , and is the largest approach set which contains  $A$ .

In view of this lemma it might possibly be more desirable, when  $G$  is not quasi-regular normal at  $z$  to define an approach set at  $z$  to be a set  $A$  such that there exist numbers  $\lambda_0, \lambda$ , not both zero, for which (if we write  $H(z, r) = \lambda_0 F(z, r) - \lambda G(z, r)$ )

$$H_i(z, p) = H_i(z, r), \quad i = 0, 1, \dots, q,$$

for all  $p, r$  in  $A$ . The interest in this remark is rather slight, since the methods of proof which are the central feature of this paper seem not to apply unless  $G(z, r)$  is quasi-regular normal.

**3. Existence of minimizing polygons.** We now add some further hypotheses concerning the integrands. These added requirements are much more stringent for problems in non-parametric form than for those in parametric form.

[(3.1) For every bounded set  $S_0$  there exist positive numbers  $\delta$  and  $a$  and a number  $b \geq 0$  such that

$$|F_i(z, z')| \leq a |z'| + bF(z_0, z')$$

wherever  $z_0$  is in  $S_0$  and  $|z - z_0| < \delta$ , and  $z'^0 > 0$ ; and  $G$  also satisfies this condition.

(3.2) On every bounded set  $S_0$  the relations

$$\lim_{|y'| \rightarrow \infty} f(x, y, y')/|y'| = \infty, \quad \lim_{|y'| \rightarrow \infty} g(x, y, y')/f(x, y, y') = 0$$

hold uniformly in  $(x, y)$ .

(3.3) The points  $z_1 = (x_0, y_0)$  and  $z_2 = (X, Y)$  and the integrals  $\mathcal{F}, \mathcal{G}$  have the property that there are constants  $a_0 \geq 0$  and  $a_1$  such that for every number  $H$  there is a bounded set  $S_H$  containing all the a.c. curves  $C: y = y(x)$  joining  $z_1$  and  $z_2$  and having  $a_0 \mathcal{F}(C) + a_1 \mathcal{G}(C) < H$ .]

For problems in parametric form the role of these hypotheses is filled by the following assumption:

(3.4) The points  $z_1, z_2$  and the integrals  $\mathcal{F}, \mathcal{G}$  have the property that if  $K$  is a class of rectifiable curves joining  $z_1$  and  $z_2$  for which  $\mathcal{F}(C)$  is bounded above and  $\mathcal{G}(C)$  is bounded, then the curves of  $K$  have uniformly bounded lengths.

Then we have the following lemma:

LEMMA 3.† Let  $v$  be an integer and  $l$  a number. If the subclass of  $K[G=l]$ , consisting of polygons of not more than  $v$  vertices, is not empty, it contains a polygon which minimizes  $\mathcal{F}(C)$  on that class of polygons.

† For problems in parametric form this is a trivial extension of Lemma 1 of III.

[Let  $y = y_n(x)$ , ( $x_0 \leq x \leq X$ ), be a sequence of polygons of not more than  $\nu$  vertices such that  $G[y_n] = l$  and  $\mathcal{Y}[y_n]$  tends to its lower bound  $\mu$ . By Lemma 1 of II, all the curves  $y = y_n(x)$  are in a circle  $Q$ . By Lemma 4 of II, the functions  $y_n(x)$  are equi-absolutely continuous; so by Ascoli's theorem there is a subsequence (which we may suppose to be the original sequence) converging uniformly to a limit function  $y = y_0(x)$ . We see readily that this limit curve is also a polygon of not more than  $\nu$  vertices, and that  $y'_n(x) \rightarrow y'_0(x)$  except at the vertices of  $y = y_0(x)$ . So by Lemma 5 of II we have  $G[y_0] = l$  and  $\mathcal{Y}[y_0] \leq \lim \mathcal{Y}[y_n] = \mu$ . But by the definition of  $\mu$  we cannot have  $\mathcal{Y}[y_0] < \mu$ ; so  $\mathcal{Y}[y_0] = \mu$ , and  $y = y_0(x)$  is the polygon sought.]

4. **Choice of a particular minimizing sequence.** Continuing with the hypotheses of the preceding section, we assume that  $l$  is a number such that  $K[G=l]$  is not empty, and we denote by  $\mu$  the greatest lower bound of  $\mathcal{Y}(C)$  on the class  $K[G=l]$ . There is [by Lemma 2 of II] a sequence of polygons  $\bar{\Pi}_n$  [of the form  $y = \bar{y}_n(x)$ ] joining  $(x_0, y_0)$  to  $(X, Y)$  and such that  $\lim \mathcal{Y}(\bar{\Pi}_n) = \mu$ ,  $\lim G(\bar{\Pi}_n) = l$ . Let  $\nu_n$  be the number of vertices of  $\bar{\Pi}_n$ . By Lemma 3 of the preceding section, there is, for each  $n$ , a polygon  $\Pi_n$  [of the form  $y = y_n(x)$ ] in  $K$  such that

$$(4.1) \quad G(\Pi_n) = G(\bar{\Pi}_n) \rightarrow l,$$

while  $\Pi_n$  minimizes  $\mathcal{Y}(C)$  on the class of polygons in  $K$  having not more than  $\nu_n$  vertices and satisfying the equation  $G(C) = G(\bar{\Pi}_n)$ . Hence

$$(4.2) \quad \mathcal{Y}(\Pi_n) \leq \mathcal{Y}(\bar{\Pi}_n).$$

From this sequence we now select a subsequence  $\bar{\Pi}_{n_i}$  such that

$$(4.3) \quad \lim_{i \rightarrow \infty} \mathcal{Y}(\bar{\Pi}_{n_i}) = \liminf_{n \rightarrow \infty} \mathcal{Y}(\Pi_n) \leq \lim_{n \rightarrow \infty} \mathcal{Y}(\bar{\Pi}_n) = \mu.$$

We may assume that this subsequence is the whole sequence, so that

$$(4.4) \quad \lim \mathcal{Y}(\Pi_n) \text{ exists, } \lim \mathcal{Y}(\Pi_n) \leq \mu.$$

By Lemma\* 1 of II the polygons  $\Pi_n$  are all in a circle  $Q$ . [So by Lemma 4 of II the functions  $y_n(x)$  are equi-absolutely continuous, and we can select a subsequence (which we suppose to be the whole sequence) converging uniformly to a limit function  $y_0(x)$ :

$$(4.5) \quad \lim_{n \rightarrow \infty} y_n(x) = y_0(x) \text{ uniformly for } x_0 \leq x \leq X.$$

This selection of a subsequence leaves (4.4) still valid.

\* Although this lemma was stated for non-parametric problems, its extension to problems in parametric form is trivial.

Hypothesis (3.2) implies that  $f(x, y, y')$  is bounded below for  $(x, y)$  in  $Q$ , say  $f(x, y, y') \geq \kappa$ . Since addition of a constant to  $f$  merely changes  $\mathcal{F}[y]$  by the same number for all curves  $y = y(x)$  of  $K$  and leaves all minima unaffected, we may assume without loss of generality that for all  $(x, y)$  in  $Q$  and all  $y'$

$$(4.6) \quad f(x, y, y') \geq 0.$$

Now we return to the parametric notation.] We have a sequence of polygons  $\{\Pi_n\}$  [converging to a limit curve  $C_0$ ] in  $K$ , and

$$(4.7) \quad \lim G(\Pi_n) = l, \quad \bar{\mu} = \lim \mathcal{F}(\Pi_n) \leq \mu.$$

Also, the lengths of the  $\Pi_n$  are bounded (by (3.4) if the problem is in parametric form, by Lemma 4 of II otherwise), and are not less than  $|z_1 - z_2|$ :

$$(4.8) \quad |z_1 - z_2| \leq \mathcal{L}(\Pi_n) \leq L.$$

On each  $\Pi_n$  we now introduce the parameter  $t = s/\mathcal{L}(\Pi_n)$ , ( $0 \leq s \leq \mathcal{L}(\Pi_n)$ ). Thus  $\Pi_n$  has the representation

$$\Pi_n: z = z_n(t), \quad 0 \leq t \leq 1,$$

with

$$(4.9) \quad 0 < |z_1 - z_2| \leq |\dot{z}_n(t)| \leq \mathcal{L}(\Pi_n) \leq L$$

except at vertices of  $\Pi_n$ .

Since the functions  $z_n$  all satisfy the same Lipschitz condition, there is a subsequence (we suppose it the whole sequence) which converges uniformly to a limit function  $z_0(t)$  on  $[0, 1]$ :

$$(4.10) \quad \lim_{n \rightarrow \infty} z_n(t) = z_0(t) \quad \text{uniformly on } [0, 1].$$

The curve  $z = z_0(t)$  is therefore a limit curve\* of the  $\Pi_n$ . [But these have the unique limit  $C_0$ ; so  $z = z_0(t)$ , ( $0 \leq t \leq 1$ ), is a representation of  $C_0$ .]

Now we define

$$(4.11) \quad \phi_n(t) = \int_0^t F(z_n(t), \dot{z}_n(t)) dt, \quad n = 0, 1, \dots$$

[By (4.6), these functions are monotonic increasing; and they are bounded, for

$$0 = \phi_n(0) \leq \phi_n(t) \leq \phi_n(1) = \int_0^1 F(z_n, \dot{z}_n) dt \rightarrow \bar{\mu} \leq \mu,$$

\* If the reader is omitting bracketed statements, we here define  $C_0$  to be the curve  $z = z_0(t)$ , ( $0 \leq t \leq 1$ ).

by (4.7).] So [by Helly's theorem] we can select a subsequence of  $\{\phi_n\}$  (we suppose it the whole sequence) which converges\* for all  $t$  in  $[0, 1]$  to a limit function  $\phi(t)$ :

$$(4.12) \quad \lim_{n \rightarrow \infty} \phi_n(t) = \phi(t), \quad 0 \leq t \leq 1.$$

For this subsequence, (4.7) and (4.10) remain valid. By Lemma 3 of II, the functions

$$(4.13) \quad \gamma_n(t) \equiv \int_0^t G(z_n(t), \dot{z}_n(t)) dt, \quad 0 \leq t \leq 1, \quad n = 0, 1, \dots,$$

are equi-absolutely continuous. Hence we can select a subsequence (we suppose it to be the whole sequence) which converges uniformly on  $[0, 1]$  to a limit function  $\gamma(t)$ :

$$(4.14) \quad \lim_{n \rightarrow \infty} \gamma_n(t) = \gamma(t) \quad \text{uniformly on } [0, 1].$$

For this subsequence (4.7), (4.10), and (4.12) remain valid.

Our task is now to prove

$$(4.15) \quad \gamma(1) = \gamma_0(1), \quad \phi(1) \geq \phi_0(1).$$

For these are the same as

$$(4.16) \quad \lim G(\Pi_n) = G(C_0), \quad \lim \mathcal{J}(\Pi_n) \geq \mathcal{J}(C_0).$$

If we can establish them, then the first equation shows that  $C_0$  is in  $K[G=l]$ , by (4.1). Hence  $\mathcal{J}(C_0) \geq \mu$ . But the second part of (4.16), with (4.7), gives  $\mathcal{J}(C_0) \leq \mu$ ; hence  $\mathcal{J}(C_0) = \mu$ , and  $y = y_0(x)$  is the curve sought.

It is worth noticing that all the properties of the sequence which have been established here are also possessed by every subsequence of this sequence.

**5. Convergence of directions at one point.** We observe that,  $\phi(t)$ ,  $\phi_0(t)$ , [being monotonic increasing, and]  $\gamma(t)$ ,  $\gamma_0(t)$ , being a.c., there is a set  $E$  of measure 1 contained in the open interval  $0 < t < 1$  and having the following property:

(5.1) *The derivatives  $\phi'(t)$ ,  $\phi_0'(t)$ ,  $\gamma'(t)$ , and  $\gamma_0'(t)$  exist and are finite for all  $t$  in  $E$ .*

We now wish to prove the following lemma:

\* For problems in parametric form we note that the integrands in (4.11) are uniformly bounded. So the  $\phi_n(t)$  all satisfy the same Lipschitz condition, and a subsequence can be selected which converges uniformly to a limit  $\phi_0(t)$ . A like remark applies to (4.14).

LEMMA 4. If  $t_1$  belongs to  $E$ , there is a subsequence  $\{\Pi_m\}$ , ( $m = n_1, n_2, \dots$ ), of  $\{\Pi_n\}$  and a sequence  $\{t_m\}$  of points of  $[0, 1]$  such that  $\lim_{m \rightarrow \infty} t_m = t_1$ , and  $z'_m(t_m)$  converges\* to a limit  $p$  [and  $p^0 > 0$ ].

[Let  $\epsilon_i$  be a sequence of positive numbers less than  $1 - t_1$ . The numbers  $[\phi(t_1 + \epsilon_i) - \phi(t_1)]/\epsilon_i$  tend to  $\phi'(t_1)$ , which is finite. Hence there is an  $h$  such that

$$|\phi(t_1 + \epsilon_i) - \phi(t_1)|/\epsilon_i < h.$$

Therefore, by (4.14), the inequality

$$(5.2) \quad |\phi_n(t_1 + \epsilon_i) - \phi_n(t_1)|/\epsilon_i < h + 1$$

holds for all  $n$  greater than a certain  $n(i)$ . We choose a sequence  $\{n_i\}$  such that  $n_1 < n_2 < \dots$  and  $n_i > n(i)$ . Since  $\phi_n$  is a.c. and (by 4.6) monotonic increasing, from (5.2) we see that there is a set of positive measure in  $[t_1, t_1 + \epsilon_i]$  on which  $|\phi'_{n_i}(t)| < h + 1$ . In this set we choose a  $t_{n_i}$  which does not define a vertex of  $\Pi_{n_i}$ ; and we use the letter  $m$  to replace  $n_i$ . Then  $z'_m(t_m)$  is defined, and

$$F(z_m(t_m), z'_m(t_m)) = \phi'_m(t_m) < h + 1.$$

Therefore by (4.9)

$$(5.3) \quad F(z_m(t_m), z'_m(t_m))/|z'_m(t_m)| < (h + 1)/|z'_m(t_m)| \leq (h + 1)/|z_1 - z_2|.$$

Hypothesis (3.2), in its parametric form, states that there is a  $\delta > 0$  such that

$$(5.4) \quad F(z, z') > \frac{h + 1}{|z_1 - z_2|}, \quad z \text{ in } Q, \quad |z'| = 1, \quad z^{0'} < \delta.$$

Therefore from (5.3) we conclude

$$(5.5) \quad z_m^{0'}(t_m) \geq \delta.$$

From the bounded sequence  $\{z'_m(t_m)\}$  we now choose a subsequence converging to a limit  $p$ ; we may suppose this subsequence to be the whole sequence. Then, by (5.5),  $p^0 \geq \delta$ , and the lemma is established.]

**6. The directions of the sides of the polygons.** In our sketch of the proof, in the introduction, the two essentials were first to show that the directions of the sides of the  $\Pi_n$  were, on short arcs, near approach sets, and second to establish an order relation between the sides. The next lemma treats the first of these two needs.

\* For problems in parametric form this is an immediate consequence of the Bolzano-Weierstrass theorem, since the lengths  $|z'_m(t_m)|$  are bounded.

LEMMA 5. Let  $\{\Pi_n\}$  be the sequence of polygons specified in §4, and let  $t_0$  be a point of  $E$ . Then either

(i) there is a subsequence  $\{\Pi_m\}$  of  $\{\Pi_n\}$  and a sequence of numbers  $t_m \rightarrow t_0$  such that each  $t_m$  is contained in an interval  $h_m \leq t_m \leq k_m$  on which  $z_m(t)$  is linear, and  $k_m - h_m$  does not tend to 0, or

(ii) there is an approach set  $A$  at  $z_0(t_0)$  and a subsequence  $\{\Pi_m\}$  such that if  $t_m \rightarrow t_0$  and  $z_m(t_m)$  is not a vertex\* of  $\Pi_m$ , then all accumulation points of the  $z'_m(t_m)$  belong to  $A$ .

Suppose that (i) does not hold. Let  $\{t_n\}$  be a sequence tending to  $t_0$ , and let  $h_n, k_n$  define the initial and final points of the side of  $\Pi_n$  to which  $t_n$  belongs. Then  $h_n \rightarrow t_0$  and  $k_n \rightarrow t_0$ . For otherwise either  $t_n - h_n$  or  $k_n - t_n$  would not tend to 0, and  $z_n(t)$  would be linear on the intervals  $(h_n, t_n)$  and  $(t_n, k_n)$ .

By Lemma 4, there is a subsequence  $\{\Pi_h\}$  of  $\{\Pi_n\}$  and a sequence of numbers  $t_h^1$  such that  $t_h^1 \rightarrow t_0$  and  $z'_h(t_h^1)$  tends to a limit  $p_1$  [with  $p_1^0 > 0$ ]. If for every sequence  $t_h \rightarrow t_0$  the relation  $z'_h(t_h) \rightarrow p_1$  holds, then case (ii) holds. Otherwise, for some sequence  $t_h^2 \rightarrow t_0$  the vectors  $z'_h(t_h^2)$  have an accumulation point  $p_2 \neq p_1$ . For a subsequence  $\{t_w^2\}$  we have  $z'_w(t_w^2) \rightarrow p_2$ , while it is still true that  $z'_w(t_w^1) \rightarrow p_1$ .

To be specific, we suppose  $t_w^1 < t_w^2$  for all  $w$ . Let  $v_0, v_1, \dots, v_{s+1}$  be values of  $t$  which define successive vertices of  $\Pi_w$  and satisfy  $v_0 < t_w^1 < v_1 < \dots < v_s < t_w^2 < v_{s+1}$ . By the first paragraph,  $v_0$  and  $v_{s+1}$  tend to  $t_0$  as  $w \rightarrow \infty$ . If we displace the vertices  $z_w(v_1), \dots, z_w(v_s)$  all by the same amount  $(0, 0, \dots, 0, \tau, 0, \dots, 0)$ , where  $\tau$  is in the  $(j+1)$ st place, we obtain a polygon  $\Pi_w^j(\tau)$ . [For small  $\tau$  this can be represented in the form  $z^i = z^i(z^0)$ ,  $(i=1, \dots, n)$ .] From formula (2.8) of I we have

$$\begin{aligned}
 (6.1) \quad \frac{d}{d\tau} \mathcal{F}(\Pi_w^j(\tau)) \Big|_{\tau=0} &= F_j(z_w(t), z'_w(t)) - F_j(z_w(\bar{t}), z'_w(\bar{t})) \\
 &+ \int_{v_0}^{v_1} \frac{t - v_0}{v_1 - v_0} F_{ji}(z_w, z'_w) dt \\
 &+ \int_{v_s}^{v_{s+1}} \frac{t - v_{s+1}}{v_s - v_{s+1}} F_{ji}(z_w, z'_w) dt \\
 &+ \int_{v_1}^{v_s} F_{ji}(z_w, z'_w) dt,
 \end{aligned}$$

where  $v_0 < \bar{t} < v_1$  and  $v_s < \bar{t} < v_{s+1}$ .

Let us denote the sum of the last three terms in (5.6) by  $\theta_j$ . [By (3.1) and (4.9) the inequality

\* This requirement is not essential.

$$(6.2) \quad \begin{aligned} |\theta_i| &\leq \int_{v_0}^{v_{s+1}} \{a|\dot{z}_w| + bF(z_w, \dot{z}_w)\} dt \\ &\leq aL|v_{s+1} - v_0| + b\{\phi_w(v_{s+1}) - \phi_w(v_0)\} \end{aligned}$$

holds.\*] Also, on the intervals  $(v_0, v_1)$  and  $(v_s, v_{s+1})$  the derivative  $z'_w(t)$  is constant. Hence (5.1) can be written

$$(6.3) \quad D_i \equiv \frac{d}{dt} \mathcal{F}(\Pi_w(\tau)) \Big|_{\tau=0} = F_i(z_w(\bar{t}), z'_w(t_w^2)) - F_i(z_w(\bar{t}), z'_w(t_w^1)) + \theta_i.$$

A similar result holds for  $G(C)$ :

$$(6.4) \quad \bar{D}_i \equiv \frac{d}{d\tau} \mathcal{G}(\Pi_w(\tau)) \Big|_{\tau=0} = G_i(z_w(\bar{t}), z'_w(t_w^2)) - G_i(z_w(\bar{t}), z'_w(t_w^1)) + \bar{\theta}_i, \\ v_0 < \bar{t} < v_1, \quad v_s < \bar{t} < v_{s+1},$$

where for non-parametric problems

$$(6.5) \quad |\bar{\theta}_i| \leq \bar{a}L(v_{s+1} - v_0) + \bar{b}\{\gamma_w(v_{s+1}) - \gamma_w(v_0)\},$$

while for parametric problems

$$(6.5') \quad |\bar{\theta}_i| \leq \bar{M}(v_{s+1} - v_0).$$

Now we show that the matrix

$$(6.6) \quad \begin{pmatrix} D_0 & D_1 & \cdots & D_q \\ \bar{D}_0 & \bar{D}_1 & \cdots & \bar{D}_q \end{pmatrix}$$

is of rank less than 2. Otherwise, suppose, say, that the determinant

$$(6.7) \quad \begin{vmatrix} D_j & D_k \\ \bar{D}_j & \bar{D}_k \end{vmatrix}$$

is not 0. We form the polygon  $\Pi_w(\tau, \sigma)$  by displacing each of the vertices  $z_w(v_1), \dots, z_w(v_s)$  by the amount  $(0, 0, \dots, \tau, 0, \dots, 0, \sigma, \dots, 0)$ , where  $\tau$  is in the  $(j+1)$ st place and  $\sigma$  in the  $(k+1)$ st. Then the jacobian of the functions  $\mathcal{F}(\Pi_w(\tau, \sigma)), \mathcal{G}(\Pi_w(\tau, \sigma))$  at  $\tau=\sigma=0$  is exactly the determinant (6.7). The equations

$$(6.8) \quad \begin{aligned} \mathcal{F}(\Pi_w(\tau, \sigma)) + u - \mathcal{F}(\Pi_w) &= 0, \\ \mathcal{G}(\Pi_w(\tau, \sigma)) - \mathcal{G}(\Pi_w) &= 0 \end{aligned}$$

have the initial solutions  $\tau=\sigma=u=0$ . By the implicit functions theorem, we can solve for  $\tau, \sigma$  as functions of  $u$ , defined for all  $u$  near 0. However, if  $u > 0$

\* For parametric problems this is replaced by the inequality (6.2')  $|\theta_i| \leq M(v_{s+1} - v_0)$ , where  $M$  is an upper bound for  $|F_w(z, z')|$  for  $z$  near  $C_0$  and  $|z'| \leq L$ .

these functions would define values of  $\tau, \sigma$  for which (6.8) held; that is,

$$(6.9) \quad \begin{aligned} G(\Pi_w(\tau, \sigma)) &= G(\Pi_w), \\ \mathcal{J}(\Pi_w(\tau, \sigma)) &< \mathcal{J}(\Pi_w). \end{aligned}$$

This is incompatible with the minimizing property of the polygons  $\Pi_w$ . Hence the matrix (6.6) is of rank less than two.

We now write (6.6) in the form

$$(6.10) \quad \begin{pmatrix} F_i(z_w(\bar{t}_w), z_w'(t_w^2)) - F_i(z_w(\bar{t}_w), z_w'(t_w^1)) + \theta_i \\ G_j(z_w(\bar{t}_w), z_w'(t_w^2)) - G_j(z_w(\bar{t}_w), z_w'(t_w^1)) + \bar{\theta}_j \end{pmatrix}.$$

For parametric problems inequalities (6.2') and (6.5') make it evident that  $\theta_i$  and  $\bar{\theta}_j$  approach 0 with  $v_{s+1} - v_0$ . But  $v_{s+1}$  and  $v_0$  approach  $t_0$  as  $w \rightarrow \infty$ ; so  $\theta_i$  and  $\bar{\theta}_j$  approach 0 as  $w \rightarrow \infty$ . [For problems in non-parametric form we use (6.2) and (6.5). Since  $\phi'(t_0)$  exists, and  $\phi(t)$  is monotonic increasing, for every positive  $\epsilon$  there is a  $\delta > 0$  such that

$$(6.11) \quad \phi(t_0) - \epsilon < \phi(t_0 - \delta) \leq \phi(t_0) \leq \phi(t_0 + \delta) < \phi(t_0) + \epsilon;$$

and if  $w$  is large enough,

$$|\phi_w(t_0 - \delta) - \phi(t_0 - \delta)| < \epsilon, \quad |\phi_w(t_0 + \delta) - \phi(t_0 + \delta)| < \epsilon.$$

Hence if

$$(6.12) \quad t_0 - \delta < t < t_0 + \delta,$$

then

$$(6.13) \quad \phi(t_0) - 2\epsilon < \phi_w(t_0 - \delta) \leq \phi_w(t) \leq \phi_w(t_0 + \delta) < \phi(t_0) + 2\epsilon.$$

For all large  $w$  the points  $v_0$  and  $v_{s+1}$  both satisfy (6.12), since they both approach  $t_0$  as  $w \rightarrow \infty$ ; so by (6.13) we have

$$(6.14) \quad |\phi_w(v_{s+1}) - \phi_w(v_0)| < 4\epsilon.$$

Thus  $\phi_w(v_{s+1}) - \phi_w(v_0)$  approaches 0 as  $w \rightarrow \infty$ . The other term  $aL(v_{s+1} - v_0)$  also tends to 0, as previously observed; so  $\theta_i$  tends to zero with  $1/w$ . The proof that  $\bar{\theta}_j \rightarrow 0$  is easier. The functions  $\gamma_w(t)$  are equi-absolutely continuous; so when  $w \rightarrow \infty$  (and consequently  $v_{s+1} - v_0 \rightarrow 0$ ) the term  $\gamma_w(v_{s+1}) - \gamma_w(v_0)$  tends to 0. The other term in (5.10) also tends to 0; hence  $\bar{\theta}_j$  approaches 0 as  $w \rightarrow \infty$ .

Since  $z_w(t)$  tends uniformly to  $z_0(t)$ , and the numbers  $\bar{t}_w, \bar{t}_w, \bar{t}_w$ , and  $\bar{t}_w$  all tend to  $t_0$ , we have

$$(6.15) \quad \lim z_w(\bar{t}_w) = \lim z_w(\bar{t}_w) = \lim z_w(\bar{t}_w) = \lim z_w(\bar{t}_w) = z_0(t_0).$$

Finally, we already know that  $\lim z_w'(t_w^1) = p_1$  [where  $p_1^0 > 0$ ], and that

$\lim z'_w(t_w^2) = p_2$ . Therefore [from (2.4) it follows that  $p_2^0 > 0$ , and by Lemma 1] (for parametric problems by Lemma 1') the vectors  $p_1$  and  $p_2$  belong to an approach set at  $z_0(t_0)$ , and the matrix

$$(6.16) \quad \begin{pmatrix} F_j(z_0(t_0), p_h) & F_j(z_0(t_0), p_k) \\ G_j(z_0(t_0), p_h) & G_j(z_0(t_0), p_k) \end{pmatrix}$$

has rank less than two, where  $h=1$  and  $k=2$ .

Now we can define the approach set  $A$  to be the maximal approach set at  $z$  containing  $p_1$  and  $p_2$  (see remark at end of §2).

Suppose now that there is a third sequence  $t_w^3 \rightarrow t_0$  such that  $z'_w(t_w^3)$  has an accumulation point  $p_3$  different from  $p_1$  and  $p_2$ . We select a subsequence  $\{l\}$  of the integers  $\{w\}$  such that  $z'_l(t_l^3) \rightarrow p_3$ . By the preceding argument, the matrix (6.16) has rank less than two if  $h=1$  and  $k=3$  or if  $h=2$  and  $k=3$ . So  $p_3$  belongs to an approach set at  $z$  containing  $p_1$  and  $p_2$ , and therefore belongs to the maximal set  $A$ . The lemma is thus established.

**7. A lemma on the order of sides.** The next lemma is the one by which the  $\Omega$ -function is made useful to us.

**LEMMA 6.** *Let  $p^*, r^*$  be distinct unit vectors [with  $p^{*0} > 0$  and  $r^{*0} > 0$ ] belonging to an approach set  $A$  at  $z^*$ . Let the inequality*

$$\Omega_H(z^*, p^*, r^*) < 0$$

*hold, where  $H(z, z') = H(z, z'; \lambda(z^*, A)) = F(z, z') - \lambda(z^*, A)G(z, z')$ . Let  $G(z, r)$  be quasi-regular normal (either positive or negative) at  $z^*$ . Then there is a  $\delta > 0$  such that if  $ABC$  is a polygon of two sides lying entirely in the  $\delta$ -neighborhood of  $z^*$ , the side  $AB$  having direction  $r$  such that  $|r - r^*| < \delta$  and the side  $BC$  having direction  $p$  such that  $|p - p^*| < \delta$ , then  $ABC$  does not minimize  $\mathcal{F}(C)$  in the class of all two-sided polygons  $ADC$  joining  $A$  and  $C$  and having  $G(ADC) = G(ABC)$ .*

To be specific we assume that  $G(z, r)$  is positive quasi-regular normal at  $z^*$ . This involves no loss of generality. For if  $G(z, r)$  is negative quasi-regular normal at  $z^*$  we have only to replace it by  $-G(z, r)$ . This causes  $\lambda(z^*, A)$  to be replaced by  $-\lambda(z^*, A)$ , and  $H(z, r)$  is unchanged.

In order to avoid repetition we assume that every positive constant introduced in this proof is less than one, and we let  $M$  be an upper bound for the sums

$$2 \sum_{i=0}^q \{ |H_{ri}(z, r)| + |G_{ri}(z, r)| \}$$

for all  $(z, r)$  such that  $|z - z^*| \leq 5$  and  $|r| = 1$  [and  $r^0 \geq \frac{1}{2} \min(p^{*0}, r^{*0})$ ]. We shall connote that a vector is a unit vector by giving it a subscript  $u$ .

By hypothesis,  $\Omega_H(z^*, p^*, r^*) < -3m < 0$ . Thus by continuity there are positive numbers  $\zeta, \delta$  (we take  $\zeta, \delta, m$  all less than 1) such that

$$(7.1) \quad \Omega_H(z, p_u, r_u) < -2m, \quad [p_u^0 > \frac{1}{2} \min(p^{*0}, r^{*0}), \quad r_u^0 > \frac{1}{2} \min(p^{*0}, r^{*0})]$$

if  $|z - z^*| < 5\zeta$ ,  $|p_u - p^*| < 3\delta$ , and  $|r_u - r^*| < 3\delta$ .

Again by hypothesis, the numbers  $\mathcal{E}_G(z^*, p^*, r^*)$  and  $\mathcal{E}_G(z^*, r^*, p^*)$  are positive. Let  $3e$ , ( $0 < e < 1$ ), be smaller than the smaller of them. Then if  $\zeta$  and  $\delta$  are small enough,

$$(7.2) \quad G(z, r_u) - r_u^a G_a(\bar{z}, p_u) > 2e, \quad G(z, p_u) - p_u^a G_a(\bar{z}, r_u) > 2e$$

if  $z$  and  $\bar{z}$  are in the  $5\zeta$ -neighborhood of  $z^*$  and  $|p_u - p^*| < 3\delta$ ,  $|r_u - r^*| < 3\delta$ . By definition of  $M$ ,

$$(7.3) \quad |\Omega_G(z, p_u, r_u)| < M \text{ if } |z - z^*| < 5\zeta, \quad |p_u - p^*| < 3\delta, \quad |r_u - r^*| < 3\delta.$$

By Lemma 2,  $H_i(z^*, p^*) = H_i(z^*, r^*)$ , ( $i=0, 1, \dots, q$ ). If we multiply by  $p^{*i}$  or by  $r^{*i}$  and sum, we find

$$(7.4) \quad \mathcal{E}_H(z^*, p^*, r^*) = \mathcal{E}_H(z^*, r^*, p^*) = 0.$$

Then if  $\zeta$  and  $\delta$  are small enough, we find, as in establishing (7.2), that

$$(7.5) \quad \begin{aligned} |H(z, r_u) - r_u^a H_a(\bar{z}, p_u)| &< em/M, \\ |H(z, p_u) - p_u^a H_a(\bar{z}, r_u)| &< em/M \end{aligned}$$

if  $|z - z^*| < 5\zeta$ ,  $|\bar{z} - z^*| < 5\zeta$ ,  $|p_u - p^*| < 3\delta$ , and  $|r_u - r^*| < 3\delta$ .

Let  $ABC$  be a two-sided polygon lying in the  $\zeta$ -neighborhood of  $z^*$ . We denote  $A, B, C$  by  $z_1, z_2, z_3$ , respectively, and for their directions and lengths we use the symbols

$$(7.6) \quad \begin{aligned} p_1 &= (z_2 - z_1)/|z_2 - z_1|, & p_2 &= (z_3 - z_2)/|z_3 - z_2|, \\ l_1 &= |z_2 - z_1|, & l_2 &= |z_3 - z_2|; \end{aligned}$$

and we assume

$$(7.7) \quad |p_1 - r^*| < \delta, \quad |p_2 - p^*| < \delta.$$

We now distinguish the two cases  $l_1 \geq l_2$ ,  $l_1 < l_2$ . In the former case we write

$$(7.8) \quad z_r = z_1 + (1 + \tau)(z_3 - z_2);$$

in the latter case,

$$(7.9) \quad z_r = z_3 + (1 + \tau)(z_1 - z_2).$$

The two cases differ only trivially; so we suppose, to be specific, that  $l_1 \geq l_2$ ; the alterations needed to cover the other case are obvious. If  $|\tau| < 1$ , then  $|z_r - z_1| < 2|z_3 - z_2| = 2l_2 < 4\zeta$ ; so  $z_r$  is in the  $5\zeta$ -neighborhood of  $z^*$ . Therefore

for  $|\tau| < 1$  the quadrilateral  $z_1 z_2 z_3 z_r$  lies entirely in the sphere  $|z - z^*| < 5\delta$ .

By an elementary trigonometric computation we find that if  $|\tau| < \delta l_1/l_2$  (and therefore if  $|\tau| < \delta$ , since  $l_1 \geq l_2$ ), then the unit vector

$$(7.10) \quad p(\tau) \equiv (z_3 - z_r)/|z_3 - z_r|$$

satisfies the inequality  $|p(\tau) - p_1| < 2\delta$ . Hence, by (7.7),

$$(7.11) \quad |p(\tau) - r^*| < 3\delta \quad \text{if} \quad |\tau| < \delta.$$

Also, if  $|\tau| < 1$ , then (by (7.8))

$$(7.12) \quad |z_r - z_3| = |z_1 - z_2 + \tau(z_3 - z_2)| \leq |z_1 - z_2| + |z_3 - z_2| \leq 2l_1.$$

We henceforth assume

$$(7.13) \quad |\tau| < \delta.$$

The point  $z_r$  will also be denoted by  $D_r$ . The figure  $ABCD_0$  is a parallelogram, by (7.8). Hence by formula (3.19) of I

$$(7.14) \quad \mathcal{K}(AD_0C) - \mathcal{K}(ABC) = -l_1 l_2 \Omega_H(\bar{z}, p_1, p_2),$$

$$(7.15) \quad \mathcal{G}(AD_0C) - \mathcal{G}(ABC) = -l_1 l_2 \Omega_G(\bar{z}, p_1, p_2),$$

where  $\bar{z}$  and  $\tilde{z}$  are in the parallelogram  $ABCD_0$ . If we recall that  $\Omega_H(z, p, r) = -\Omega_H(z, r, p)$ , these inequalities, with (7.7), (7.1), and (7.3), yield

$$(7.16) \quad \mathcal{K}(AD_0C) - \mathcal{K}(ABC) < -2l_1 l_2 m,$$

$$(7.17) \quad |\mathcal{G}(AD_0C) - \mathcal{G}(ABC)| < l_1 l_2 M.$$

Now we compute the derivative of  $\mathcal{G}(AD, C)$  with respect to  $\tau$ . The side  $AD$  is defined by the equation  $z = z_1 + t(z_3 - z_2)$ , ( $0 \leq t \leq 1 + \tau$ ). So

$$(7.18) \quad \begin{aligned} \frac{d}{d\tau} \mathcal{G}(AD, C) &= \frac{d}{d\tau} \int_0^{1+\tau} G(z_1 + t(z_3 - z_2), z_3 - z_2) dt \\ &= G(z_r, z_3 - z_2) = l_2 G(z_r, p_2). \end{aligned}$$

The derivative of  $G(D, C)$  we compute by formula (2.8) of I, taking  $\pi_1 = z_3 - z_2$ ,  $\pi_2 = 0$ . The side  $D, C$  is defined by the equation  $z = (1-t)z_r + tz_3$ , ( $0 \leq t \leq 1$ ); so

$$(7.19) \quad \begin{aligned} \frac{d}{d\tau} \mathcal{G}(D, C) &= -G_\alpha(\bar{z}, z_3 - z_2)[z_3^\alpha - z_2^\alpha] \\ &\quad + \int_0^1 G_{\alpha\alpha}((1-t)z_r + tz_3, z_3 - z_r)[z_3^\alpha - z_r^\alpha](1-t)dt \\ &= -l_2 G_\alpha(\bar{z}, p_2)p_2^\alpha \\ &\quad + \int_0^1 G_{\alpha\alpha}((1-t)z_r + tz_3, p(\tau))p_2^\alpha l_2 |z_3 - z_r| (1-t)dt, \end{aligned}$$

where  $\bar{z}$  is on the line segment  $D, C$ . By the definition of  $M$ , together with (7.11) and (7.12), the integrand in (7.19) does not exceed  $2Ml_1l_2(1-t)$  in absolute value. Hence the absolute value of the integral is at most  $Ml_1l_2$ , and if we add (7.18) and (7.19) we find

$$(7.20) \quad \frac{d}{d\tau} G(AD, C) = l_2[G_\alpha(z_\tau, p_2) - p_2^\alpha G_\alpha(\bar{z}, p(\tau))] + \theta Ml_1l_2, \quad |\theta| \leq 1.$$

The points  $z_\tau$  and  $\bar{z}$  are in the  $5\zeta$ -neighborhood of  $z^*$ , and (7.11) and (7.7) hold; so by (7.2),

$$(7.21) \quad \frac{d}{d\tau} G(AD, C) > 2l_2e + \theta Ml_1l_2, \quad |\theta| \leq 1,$$

if  $|\tau| \leq \delta$ .

The conditions thus far imposed on  $\zeta$  and  $\delta$  merely require that they be sufficiently small. We now require

$$(7.22) \quad \zeta < em\delta/2M^2.$$

Then

$$(7.23) \quad l_1 < 2\zeta < c\delta m/M^2 < e/M,$$

and (7.21) yields

$$(7.24) \quad \frac{d}{d\tau} G(AD, C) > l_2e \quad \text{if} \quad |\tau| < \delta.$$

A formula similar to (7.20) holds for  $\mathfrak{H}$  also. Here, however, we can use (7.5) and (7.23) to obtain

$$(7.25) \quad \left| \frac{d}{d\tau} \mathfrak{H}(AD, C) \right| < l_2[em/M] + \theta Ml_1l_2, \quad |\theta| \leq 1, \\ < 2lem/M.$$

As  $\tau$  traverses the interval  $[-\delta, \delta]$  inequality (7.24) remains valid; so by the theorem of mean value

$$(7.26) \quad \begin{aligned} G(AD_\delta C) - G(AD_0 C) &> \delta l_2e, \\ G(AD_0 C) - G(AD_{-\delta} C) &> \delta l_2e. \end{aligned}$$

By (7.17) and (7.23)

$$(7.27) \quad |G(AD_0 C) - G(ABC)| < c\delta l_2;$$

so by (7.26) the number  $G(ABC)$  lies between  $G(AD_{-\delta} C)$  and  $G(AD_\delta C)$ . But

$G(AD, C)$  is a continuous function of  $\tau$ ; so there is a value of  $\tau$ , we call it  $\sigma$ , such that

$$(7.28) \quad G(AD_\sigma C) = G(ABC).$$

Using another form of the theorem of mean value, we obtain

$$(7.29) \quad \frac{\mathfrak{C}(AD_\sigma C) - \mathfrak{C}(AD_0 C)}{G(AD_\sigma C) - G(AD_0 C)} = \frac{\frac{d}{d\tau} \mathfrak{C}(AD_\tau C)}{\frac{d}{d\tau} G(AD_\tau C)},$$

the right-hand member being calculated for some  $\tau$  between 0 and  $\sigma$ . By (7.21) and (7.22) the right-hand member of (7.29) has an absolute value less than  $2m/M$ ; so, using (7.29), (7.28), (7.17), and (7.16), we obtain

$$(7.30) \quad \begin{aligned} \mathfrak{C}(AD_\sigma C) - \mathfrak{C}(AD_0 C) &< (2m/M) | G(AD_\sigma C) - G(AD_0 C) | \\ &= (2m/M) | G(ABC) - G(AD_0 C) | \\ &< 2ml_1 l_2 < \mathfrak{C}(ABC) - \mathfrak{C}(AD_0 C). \end{aligned}$$

That is,

$$(7.31) \quad \mathfrak{C}(AD_\sigma C) < \mathfrak{C}(ABC).$$

It follows from (7.28) and (7.31) that

$$(7.32) \quad \begin{aligned} \mathcal{F}(AD_\sigma C) &= \mathfrak{C}(AD_\sigma C) + \lambda(z^*, A) G(AD_\sigma C) \\ &< \mathfrak{C}(ABC) + \lambda(z^*, A) G(ABC) \\ &= \mathcal{F}(ABC). \end{aligned}$$

We have thus shown that if  $ABC$  is in the  $\zeta$ -neighborhood of  $z^*$  and inequalities (7.7) hold, then  $ABC$  fails to minimize  $G(ADC)$  in the class of two-sided polygons  $ADC$  such that  $G(ADC) = G(ABC)$ . For (7.28) and (7.32) show that  $AD_\sigma C$  is in the given class and gives a smaller value to  $\mathcal{F}(C)$ . If we let  $\delta_1$  be the smaller of  $\zeta$  and  $\delta$ , it then serves as the  $\delta$  of the conclusion of the lemma. The proof is therefore complete.

**8. The existence theorem.** We now make another assumption concerning  $\mathcal{F}$  and  $G$ :

(8.1) For each  $z_0$ , every approach set  $A$  at  $z_0$  consists of the positive multiples of a finite number of unit vectors  $p_1, \dots, p_k$ ; and these can be so ordered that†  $\Omega_H(z_0, p_i, p_i) < 0$  if  $i < j$ .

We wish to prove the following lemma:

† As usual,  $H(z, r) = F(z, r) - \lambda(z_0, A)G(z, r)$ .

LEMMA 7. If  $t_0$  is in  $E$ , then

$$(8.2) \quad \phi'_0(t_0) = \phi'(t_0)$$

and

$$(8.3) \quad \gamma'_0(t_0) = \gamma'(t_0),$$

where  $\phi, \phi_0, \gamma, \gamma_0$  are the functions defined in §4.

At  $t_0$ , either case (i) or case (ii) of Lemma 5 holds. Suppose case (i). Then there is a sequence  $\{\Pi_m\}$  and a sequence  $t_m \rightarrow t_0$ , each  $t_m$  being contained in an interval  $h_m < t_m < k_m$  on which  $z_m(t)$  is linear, and  $k_m - h_m$  does not tend to zero. We now select a subsequence  $\{\Pi_i\}$  for which  $k_i - h_i$  tends to a limit greater than zero, and further select a subsequence  $\{\Pi_p\}$  of  $\{\Pi_i\}$  such that  $h_p$  and  $k_p$  tend to limits  $h, k$ , respectively. Then  $k > h$ . Also, since  $h_p < t_p < k_p$  and  $t_p \rightarrow t_0$ , it is clear that  $h \leq t_0 \leq k$ . Suppose, to be specific, that  $h \leq t_0 < k$ .

We now choose numbers  $l, m$  such that  $h < l < m < k$ . On the interval  $[l, m]$  all functions  $z_p(t)$  except a finite number are linear. Since  $z_p(t)$  tends to  $z_0(t)$  and is linear on  $[l, m]$ , it is also true that  $z'_p(t)$  tends to  $z'_0(t)$ , and in fact uniformly, on  $[l, m]$ . So

$$\begin{aligned} \phi(m) - \phi(l) &= \lim_{p \rightarrow \infty} (\phi_p(m) - \phi_p(l)) = \lim_{p \rightarrow \infty} \int_l^m F(z_p, z'_p) dt \\ &= \int_l^m F(z_0, z'_0) dt = \phi_0(m) - \phi_0(l). \end{aligned}$$

Since  $\phi$  and  $\phi_0$  are continuous at  $t_0$ , we let  $l \rightarrow t_0$  and obtain

$$\phi(m) - \phi(t_0) = \phi_0(m) - \phi_0(t_0).$$

Dividing by  $m - t_0$  and letting  $m \rightarrow t_0$  gives

$$\phi'(t_0) = \phi'_0(t_0),$$

establishing (8.2). In a like manner we establish (8.3).

Now suppose that case (ii) of Lemma 5 holds. Let  $A$  be an approach set at  $z_0(t_0)$  consisting of the positive multiples of the unit vectors  $p_1, \dots, p_k$  numbered as in (8.1), and let  $\{\Pi_m\}$  be a subsequence of  $\{\Pi_n\}$  such that for every sequence  $t_m \rightarrow t_0$  all accumulation points of the sequence  $z'_m(t_m)$  are in  $A$ . Then if  $\epsilon$  is a positive number, there is a  $\delta_0 > 0$  and an  $m_0$  such that  $z'_m(t)$  is in the  $\epsilon$ -neighborhood of  $A$  if  $|t - t_0| < \delta_0$  and  $m > m_0$ . Otherwise for some  $\epsilon > 0$  we could select a subsequence  $\{\Pi_p\}$  and a sequence  $t_p \rightarrow t_0$  for which  $z'_p(t_p)$  has distance greater than or equal to  $\epsilon$  from  $A$ , so that no accumulation point of this sequence would be in  $A$ .

By Lemma 6, for each pair of numbers  $i$  and  $j > i$  there is a positive num-

ber  $\delta_{ij}$  with the properties specified in Lemma 6. Let  $\delta$  be the least of these. We suppose that the  $\epsilon$  of the preceding paragraph is so small that

$$(8.4) \quad |z'_m(t)/|z'_m(t)| - p_i| < \delta$$

for some  $i$  if  $z'_m(t)$  is in the  $\epsilon$ -neighborhood of  $A$ ; this is possible because of (4.9). Moreover, we suppose  $m_0$  so large and  $\delta_0$  so small that

$$(8.5) \quad |z_m(t) - z_0(t_0)| < \delta \quad \text{if } m > m_0, \quad |t - t_0| < \delta_0.$$

Now denote by  $t_m^1$  the least  $t$  in the interval  $(t_0 - \delta_0, t_0 + \delta_0)$  which defines a vertex of  $\Pi_m$ , and denote by  $T_m$  the greatest such  $t$ . Each side of  $\Pi_m$  between  $z_m(t_m^1)$  and  $z_m(T_m)$  has a direction  $z'_m(t)/|z'_m(t)|$  which differs from one of the  $p_i$  by less than  $\epsilon$ . By Lemma 6 and the minimizing property of each  $\Pi_n$ , no side with direction near  $p_j$  is immediately followed by one with direction near  $p_i$ , ( $i < j$ ). Hence the arc of  $\Pi_m$  between  $z_m(t_m^1)$  and  $z_m(T_m)$  consists of a subarc (possibly empty) along which  $z'_m/|z'_m|$  differs by less than  $\epsilon$  from  $p_1$ , followed by an arc along which  $z'_m/|z'_m|$  differs by less than  $\epsilon$  from  $p_2$ , followed by  $\dots$ , followed by an arc along which  $z'_m/|z'_m|$  differs from  $p_k$  by less than  $\epsilon$ . The values of  $t$  which mark the ends of these arcs we denote by  $t_m^2, t_m^3, \dots, t_m^{k+1} = T_m$ . Thus if  $t_m^j < t < t_m^{j+1}$  and  $t$  does not define a vertex of  $\Pi_m$ , the inequality

$$(8.6) \quad |z'_m(t)/|z'_m(t)| - p_j| < \epsilon$$

holds. We now select a subsequence  $\{\Pi_h\}$  such that  $t_h^j$  converges to a limit  $t^j$  as  $h \rightarrow \infty$ , ( $j=1, \dots, k+1$ ). Then  $t^1 < t_0 < t^{k+1}$ , for otherwise we would be back in case (i) of Lemma 5.

If the interval  $(t^j, t^{j+1})$ , ( $j=1, \dots, k$ ), is not empty, we choose an interval  $(l, m)$  such that  $t^j < l < m < t^{j+1}$ . Then for all large  $h$  the inequality

$$(8.7) \quad |z'_h(t)/|z'_h(t)| - p_j| < \epsilon$$

holds if  $l \leq t \leq m$  and  $z_h(t)$  is not a vertex. From (8.7) and (4.9),

$$(8.8) \quad |z'_h(t) - \mathcal{L}(\Pi_h)p_j| = \mathcal{L}(\Pi_h) \cdot |z'_h(t)/|z'_h(t)| - p_j| < \epsilon L.$$

If  $t$  and  $t+\tau$  are both in  $(l, m)$ , this yields

$$(8.9) \quad \begin{aligned} & |z_h(t+\tau) - z_h(t) - \mathcal{L}(\Pi_h)\tau p_j| \\ &= \left| \int_t^{t+\tau} \{z'_h(t) - \mathcal{L}(\Pi_h)p_j\} dt \right| \leq \epsilon L |\tau|. \end{aligned}$$

We may without loss of generality suppose that the sequence  $\{\Pi_h\}$  was so chosen that the lengths  $\mathcal{L}(\Pi_h)$  approach a limit  $L_0$ . Then, letting  $h \rightarrow \infty$ , we have

$$(8.10) \quad |z_0(t + \tau) - z_0(t) - L_0 \tau p_j| \leq \epsilon L |\tau|.$$

Now dividing by  $\tau$  and letting  $\tau \rightarrow 0$ , we obtain

$$(8.11) \quad |z'_0(t) - L_0 p_j| \leq \epsilon L,$$

provided only that  $z'_0(t)$  exists. For all large  $h$  we have  $|\mathcal{L}(\Pi_h) - L_0| < \epsilon$ , and (8.8) holds; so if  $l \leq t \leq m$  and  $z'_0(t)$  exists, then

$$(8.12) \quad \begin{aligned} |z'_h(t) - z'_0(t)| &\leq |z'_h(t) - \mathcal{L}(\Pi_h)p_j| + |\mathcal{L}(\Pi_h)p_j - L_0 p_j| \\ &\quad + |L_0 p_j - z'_0(t)| < \epsilon(2L + 1). \end{aligned}$$

Since  $F$  and  $G$  are continuous at the point  $(z_0(t_0), Lp_j)$ , for every  $\rho > 0$  we can find a neighborhood  $U$  of this point set such that

$$(8.13) \quad |F(z, z') - F(z_1, z'_1)| < \rho$$

and

$$(8.14) \quad |G(z, z') - G(z_1, z'_1)| < \rho$$

if  $(z, z')$  and  $(z_1, z'_1)$  are both in  $U$ . We now suppose that  $\delta$  and  $\epsilon$  are so small and  $h_0$  so large that we have

$$(8.15) \quad (z_h(t), z'_h(t)) \text{ and } (z_0(t), z'_0(t)) \text{ in } U$$

if  $l \leq t \leq m$  and  $h > h_0$  and the derivatives exist (see (8.8) and (8.12)). Then by (8.13), for all  $h > h_0$  we have

$$(8.16) \quad \begin{aligned} |\phi_h(m) - \phi_h(l) - [\phi_0(m) - \phi_0(l)]| &= \left| \int_l^m [F(z_h, \dot{z}_h) - F(z_0, \dot{z}_0)] dt \right| \\ &\leq \rho(m - l). \end{aligned}$$

Letting  $h \rightarrow \infty$ , we get

$$(8.17) \quad |[\phi(m) - \phi(l)] - [\phi_0(m) - \phi_0(l)]| \leq \rho(m - l).$$

Now  $\phi$  is continuous on  $[t_0 - \delta, t_0 + \delta]$ , for  $[z_h^{0'}]$  is bounded away from 0 by (8.6) (if  $\epsilon$  is smaller than  $p_j^0/L$ ), so that  $F(z_h, \dot{z}_h)$  is uniformly bounded. Also,  $\phi_0$  is a.c. by its definition. So we can let  $m \rightarrow t^{j+1}$  and  $l \rightarrow t^j$ , obtaining

$$(8.18) \quad |\{\phi(t^{j+1}) - \phi(t^j)\} - \{\phi_0(t^{j+1}) - \phi_0(t^j)\}| \leq \rho(t^{j+1} - t^j).$$

Although (8.18) was obtained under the assumption  $t^{j+1} > t^j$ , it clearly holds if  $t^{j+1} = t^j$ .

Now we add the relations (8.18) for  $j=0, 1, \dots, k$ . We find

$$(8.19) \quad |\{\phi(t^{k+1}) - \phi(t^1)\} - \{\phi_0(t^{k+1}) - \phi_0(t^1)\}| \leq \rho(t^{k+1} - t^1).$$

We can divide by  $t^{k+1} - t^1$ , since we have seen that  $t^1 < t_0 < t^{k+1}$ . But the num-

ber  $\delta_0$  could be chosen as small as desired, and  $t_0 - \delta \leq t^1 < t^{k+1} \leq t_0 + \delta_0$ . So if we let  $\delta_0$  tend to 0, the difference  $t^{k+1} - t^1$  must also approach 0, and (8.19) yields

$$(8.20) \quad |\phi'(t_0) - \phi'_0(t_0)| \leq \rho.$$

But here  $\rho$  is an arbitrary positive number; so (8.20) implies

$$(8.21) \quad \phi'(t_0) = \phi'_0(t_0).$$

Likewise we can establish (8.3).

This completes the proof of the theorem. For  $\gamma(t)$  and  $\gamma_0(t)$  are a.c., and their derivatives are equal for almost all  $t$ , and  $\gamma(0) = \gamma_0(0) = 0$ ; so  $\gamma(1) = \gamma_0(1)$ . Also,  $\phi(t)$  is monotonic increasing and  $\phi_0(t)$  is a.c., and  $\phi'(t) = \phi'_0(t)$  for almost all  $t$ ; hence

$$\phi(1) \geq \int_0^1 \dot{\phi}(t) dt = \int_0^1 \dot{\phi}_0(t) dt = \phi_0(1).$$

Thus we have established (4.15), and, as we saw in §4, this implies that  $C_0$  is the minimizing curve sought.

Collecting the hypotheses introduced at various stages of the proof, we obtain the following theorem:

**THEOREM.** *Let the functions  $f(x, y, \dot{y})$  and  $g(x, y, \dot{y})$  be defined and continuous with all their first-order partial derivatives for all  $(x, y, \dot{y})$ . Let  $F(z, z')$  and  $G(z, z')$  be the parametric integrands associated with  $f, g$ , respectively. Assume that hypotheses (2.4), (3.1), (3.2), (3.3), and (8.1) are satisfied. Let  $\mathcal{G}(C)$  be quasi-regular normal, and let  $(x_0, y_0)$  and  $(X, Y)$  be two points such that  $x_0 < X$ .*

*Then for every  $l$  the class of a.c. curves  $y = y(x)$ ,  $(x_0 \leq x \leq X)$ , joining  $(x_0, y_0)$  to  $(X, Y)$  and such that  $G(y) = l$  either is empty or contains a minimizing curve for  $\mathcal{F}(C)$ .*

**9. Example.** An example satisfying our conditions is

$$f(x, y, y') = \phi(y)(y'^2 + 1), \quad g(x, y, y') = (1 + y'^2)^{1/2},$$

where  $\phi(y) > 0$  and  $\phi'(y) > 0$ . Transforming to parametric form, we have

$$F(x, y, x', y') = \phi(y) \left( \frac{y'^2}{x'} + x' \right), \quad G(x, y, x', y') = (x'^2 + y'^2)^{1/2}.$$

To determine the approach sets at a fixed point  $(x, y)$ , we recall that, by Lemma 2, if  $(1, q)$  and  $(1, \bar{q})$  form an approach set, then for some number  $\lambda$  these same sets form an approach set for  $f - \lambda g$ . Hence if we consider the graph of  $u = f(x, y, r) - \lambda g(x, y, r)$  as a function of  $r$ , the tangent at  $r = q$  and the

tangent at  $r=\bar{q}$  coincide. (Cf. the geometric interpretation of Lemma 2.) Therefore the graph has at least two flex points between  $q$  and  $\bar{q}$ . But

$$f_{y'y'}(x, y, r) - \lambda g_{y'y'}(x, y, r) = 2\phi(y) - \lambda(1+r^2)^{-3/2}.$$

This can have at most two zeros, of opposite sign. So  $q$  and  $\bar{q}$  are of opposite sign, and no approach set can contain more than two members of the form  $(1, q_1)$  and  $(1, q_2)$ . We easily see that  $(1, q)$  and  $(1, -q)$  form an approach set. Hence given any  $(x, y, q)$ , we see that the entire approach set at  $(x, y)$  containing  $(1, q)$  consists of the positive multiples of  $(1, q)$  and  $(1, -q)$ .

We readily calculate that, independently of  $\lambda$ ,

$$\Omega_H(x, y; 1, q; 1, -q) = 2q\phi'(y)(q^2 + 1),$$

which is positive if  $q > 0$ . So (8.1) is satisfied. It remains only for us to verify (2.4). In (2.4) we may assume, if we wish, that  $|p_n| = |r_n| = 1$ , since  $F_i(z, r)$  and  $G_i(z, r)$  are positively homogeneous of degree 0 in  $r$ . Then their limits  $p, r$  are also unit vectors. With the assumption, the matrix in (2.4) is, for our example,

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

where

$$A_{11} = \phi(\bar{y}_{1,n})\{1 - (r_n^1/r_n^0)^2\} - \phi(y_{1,n})\{1 - (p_n^1/p_n^0)^2\} + \theta_{1,n},$$

$$A_{12} = 2\phi(\bar{y}_{2,n})r_n^1/r_n^0 - 2\phi(y_{2,n})p_n^1/p_n^0 + \theta_{2,n},$$

$$A_{21} = r_n^0 - p_n^0 + \bar{\theta}_{1,n}, \quad A_{22} = r_n^1 - p_n^1 + \bar{\theta}_{2,n}.$$

We must show  $\lim r_n^0 = r^0 > 0$ . Suppose the contrary. Then  $\lim |r_n^1| = 1$ , while  $\lim p_n^1 = p^1 \neq 1$  and  $\lim \bar{\theta}_{2,n} = 0$ . Therefore  $\lim A_{22} \neq 0$ . The last two terms in  $A_{11}$  and the last two in  $A_{12}$  have finite limits. If our matrix has rank less than two, then

$$A_{11}A_{22} - A_{12}A_{21} = 0, \quad n = 1, 2, \dots$$

But  $|r_n^1/r_n^0|$  tends to  $\infty$ , and in the determinant its square occurs with a coefficient which approaches a limit  $\phi(y)(r^1 - p^1)$  which is not zero, while its first power has a bounded coefficient. Thus as  $n \rightarrow \infty$  the absolute value of the determinant will also tend to  $\infty$ , contradicting the assumption that it is zero for all  $n$ . So (2.4) is established, and all the hypotheses of our theorem are satisfied.

UNIVERSITY OF VIRGINIA,  
CHARLOTTESVILLE, VA.

# SOME EXISTENCE THEOREMS IN THE CALCULUS OF VARIATIONS

## V. THE ISOPERIMETRIC PROBLEM IN PARAMETRIC FORM\*

BY

E. J. McSHANE

1. **First existence theorem.** We continue the notation of preceding papers with the trifling change that  $z$  will denote  $(z^1, \dots, z^v)$  instead of  $(z^0, \dots, z^v)$  as heretofore, and a similar change for  $z'$ . The class of all rectifiable curves joining two fixed points  $z_1, z_2$  (not necessarily distinct) will be denoted by  $K$ . The functions  $F(z, z')$  and  $G(z, z')$  are defined for all points  $z = (z^1, \dots, z^v)$  in a closed point set  $S$  and all  $z'$ . They are positively homogeneous of degree 1 in  $z'$ , are continuous for all  $z$  in  $S$  and all  $z'$ , and possess partial derivatives of first and second orders continuous except at  $z' = 0$ . We write

$$\mathcal{F}(C) = \int_C F(z, \dot{z}) dt, \quad \mathcal{G}(C) = \int_C G(z, \dot{z}) dt.$$

A hypothesis which we shall henceforth impose on our integrals is the following:

(1.1) *To each pair of numbers  $l, m$  there corresponds a number  $L$  such that if  $C$  is in  $K$  and  $|\mathcal{G}(C)| \leq l, \mathcal{F}(C) \leq m$ , then  $\mathcal{L}(C) \leq L$ .*

There are well known conditions which ensure this. For instance, if there are numbers  $a \geq 0$  and  $b$  such that  $aF + bG$  is positive definite and the set  $S$  is bounded, then (1.1) is satisfied. Or,  $S$  being unbounded, if there exist numbers  $a \geq 0$  and  $b$  for which

$$aF(z, z') + bG(z, z') \geq k |z'| (1 + |z|)^{-1},$$

$k > 0$ , then (1.1) holds.

The proofs in note IV give us, with hardly any modification, a proof of the following theorem:

---

\* Presented to the Society, April 15, 1938; received by the editors October 29, 1937 and, in revised form, February 16, 1938.

The preceding notes in this series have already appeared in these Transactions as follows: I, vol. 44 (1938), pp. 429-438; II, vol. 44 (1938), pp. 439-453; III, vol. 45 (1939), pp. 151-171; IV, vol. 45 (1939), pp. 173-196. They will be referred to by their numbers.

**THEOREM 1.** *Let the set  $S$  consist of the entire  $z$ -space, and let (1.1) hold. Let  $G(C)$  be quasi-regular normal. Suppose further that for each  $z$ , each approach set  $A$  at  $z$  contains only a finite number of unit vectors  $p_1, \dots, p_k$  which can be so ordered that  $\Omega_H(z, p_i, p_j) < 0$  if  $i < j$ , where  $H(z, r) = H(z, r; \lambda(z, A))$ . Then for every number  $l$ , either the class  $K[G=l]$  is empty or it contains a curve which minimizes  $\mathcal{F}(C)$  on the class  $K[G=l]$ .*

The proofs in IV leading up to equations (8.2) and (8.3) of IV were so designed as to apply to problems in parametric form as well as to those in non-parametric form. Hence these equations, which are  $\phi'(t) = \phi'_0(t)$  and  $\gamma'(t) = \gamma'_0(t)$  for all  $t$  in  $E$  (that is, for almost all  $t$  in  $[0, 1]$ ) are here valid. But for problems in parametric form the functions  $\phi, \phi_0, \gamma, \gamma_0$  are Lipschitzian; so this implies that  $\phi(1) = \phi_0(1)$  and  $\gamma(1) = \gamma_0(1)$ . As remarked at the end of §4 of IV, these equations imply the conclusion of the theorem.

**2. Statement of the principal existence theorem.** Before stating the next problem we introduce several definitions:

(2.1) *The point  $z_0$  of  $S$  is an ordinary point if the following conditions hold:*

- (a) *It is interior to  $S$ .*
- (b)  *$G(z, r)$  is quasi-regular normal (either positive or negative) at  $z_0$ .*
- (c) *Each approach set  $A$  at  $z_0$  contains only a finite number of unit vectors  $p_1, \dots, p_k$ , and these can be so ordered that  $\Omega_H(z_0, p_i, p_j) < 0$  if  $i < j$ ; here  $H(z, r) = F(z, r) - \lambda(z_0, A)G(z, r)$ .*

Theorem 1 required that  $S$  be the whole space and that every point  $z_0$  be ordinary. Subject to further hypotheses, our next theorem will permit  $S$  to be a closed subset of the space and will allow  $S$  to contain singular points (that is, points which are not ordinary points).

One of our hypotheses will be the following:

(2.2) *For each  $z_0$  in  $S$  there is a number  $\theta$  such that  $F(z, r) - \theta G(z, r)$  is p.q.r.† at  $z_0$ .*

We denote by  $m(z_0)$ ,  $M(z_0)$ , respectively, the greatest lower bound and the least upper bound of all numbers  $\theta$  for which  $F(z, r) - \theta G(z, r)$  is p.q.r. at  $z_0$ . It is easy to see that  $m(z) = -\infty$  if and only if  $G(z, r)$  is p.q.r. at  $z_0$ . For if  $G(z, r)$  is p.q.r. at  $z_0$  and  $\theta_0$  is any number such that  $F - \theta_0 G$  is p.q.r. at  $z_0$ , then  $F(z, r) - \theta G(z, r) = [F(z, r) - \theta_0 G(z, r)] + (\theta_0 - \theta)G(z, r)$  is also p.q.r. for all  $\theta < \theta_0$ ; so  $m(z) = -\infty$ . However, if  $G$  is not p.q.r. at  $z_0$  there are orthogonal unit vectors  $p, u$  for which  $u^\alpha G_{\alpha\beta}(z, p)u^\beta < 0$ . Then  $u^\alpha [F_{\alpha\beta}(z, p) - \theta G_{\alpha\beta}(z, p)]u^\beta$  is negative if  $\theta$  is negative and numerically large; so  $m(z_0)$  is finite. Likewise  $M(z_0) = +\infty$  if and only if  $G(z, r)$  is n.q.r. at  $z_0$ .

† Defined in (6.2) of III.

For each curve  $C: z=z(t)$ , ( $t_1 \leq t \leq t_2$ ), lying in  $S$  we define two sets of points  $t$  as follows:

(2.3)  $T_+(C)$  is the set of all points  $t$  such that  $z(t)$  is a singular point and  $G(z, r)$  is not n.q.r. at  $z(t)$ .

(2.4)  $T_-(C)$  is the set of all points  $t$  such that  $z(t)$  is a singular point and  $G(z, r)$  is not p.q.r. at  $z(t)$ .

These sets may overlap: if  $z(t)$  is a singular point and  $G(z, r)$  is neither p.q.r. nor n.q.r. at  $z(t)$ , then  $t$  belongs to both  $T_+(C)$  and  $T_-(C)$ . Also they may vanish simultaneously; this happens whenever  $G(z, r)$  is linear in the variables  $r^i$  at each singular point  $z(t)$ .

Our next definition is as follows:

(2.5) If  $C: z=z(t)$  is a curve lying in  $S$ , then if  $T_+(C)$  is not empty, we define  $M(C)$  to be the greatest lower bound of  $M(z(t))$  for all  $t$  in  $T_+(C)$ ; if  $T_-(C)$  is not empty, we define  $m(C)$  to be the least upper bound of  $m(z(t))$  for all  $t$  in  $T_-(C)$ .

If  $M(C)$  is defined, it is not  $+\infty$ . For if  $M(C)$  is defined, the set  $T_+(C)$  is not empty. Let  $t_0$  be a point in it. At  $z(t_0)$  the function  $G(z, r)$  is not n.q.r.; so  $M(z(t_0))$  is not  $+\infty$ . Hence  $M(C) \leq M(z(t_0)) < \infty$ . Likewise, if  $m(C)$  is defined, it is not  $-\infty$ . It is interesting to observe that if  $G(z(t), r)$  is positive regular for all  $t$  and if  $C$  contains singular points  $z(t)$ , then  $M(C)$  is defined and finite, while  $m(C)$  is undefined. Likewise, if  $G(z(t), r)$  is negative regular for all  $t$  and if  $C$  contains singular points, then  $m(C)$  is defined and finite, while  $M(C)$  is undefined. We prove the first statement; the proof of the second is similar. The set  $T_+(C)$  here consists of all singular points, and by hypothesis is not empty; so  $M(C)$  is defined. As always,  $M(C) < \infty$ . The quadratic form  $u^a G_{ab}(z(t), p) u^b$  is positive for all  $t$  and all pairs of orthogonal unit vectors  $u, p$ . Let  $\nu$  be its greatest lower bound; then  $\nu > 0$ .

The quadratic form  $u^a F_{ab}(z(t), p) u^b$  is bounded, say by  $N$ , in absolute value, for the same arguments. Then  $F(z(t), r) + (N/\nu)G(z(t), r)$  is p.q.r. for all  $t$ ; so  $M(z(t)) \geq -N/\nu$  for each  $t$ , and  $M(C) \geq -N/\nu > -\infty$ . Therefore  $M(C)$  is finite.

We now state our principal theorem.

**THEOREM 2.** *Let the following hypotheses be satisfied:*

(a)  $S$  is closed.

(b) Hypotheses (1.1) and (2.2) hold.

(c) For every curve  $C$  of  $K$ , either  $T_+(C)$  is empty or there exists† a curve  $\Gamma^*: z=\zeta(\tau)$ , ( $0 \leq \tau \leq \epsilon$ ), with the properties:

† This implies that  $M(C)$  is finite. For  $M(C)$  is never  $+\infty$ , and if  $M(C)$  were  $-\infty$ , inequalities (2.6) and (2.7) could not hold. Likewise, from (d) we conclude that  $m(C)$  is undefined or is finite.

- (i) The length of  $\Gamma^*$  is not zero.
- (ii)  $\xi(0)$  is on  $C$ .
- (iii) For almost all  $\tau$  in  $[0, \epsilon]$  the inequalities

$$(2.6) \quad G(\xi(\tau), \xi(\tau)) + G(\xi(\tau), -\xi(\tau)) > 0,$$

$$(2.7) \quad F(\xi(\tau), \xi(\tau)) + F(\xi(\tau), -\xi(\tau)) \leq M(C)[G(\xi(\tau), \xi(\tau)) + G(\xi(\tau), -\xi(\tau))]$$

hold.

(d) For every curve  $C$  of  $K$ , either  $T_-(C)$  is empty or there exists a curve  $\Gamma^*: z = \xi(\tau)$ ,  $(0 \leq \tau \leq \epsilon)$ , with the properties:

- (i) The length of  $\Gamma^*$  is not zero.
- (ii)  $\xi(0)$  is on  $C$ .
- (iii) For almost all  $\tau$  in  $[0, \epsilon]$  the inequalities

$$(2.8) \quad G(\xi(\tau), \xi(\tau)) + G(\xi(\tau), -\xi(\tau)) < 0,$$

$$(2.9) \quad F(\xi(\tau), \xi(\tau)) + F(\xi(\tau), -\xi(\tau)) \leq m(C)[G(\xi(\tau), \xi(\tau)) + G(\xi(\tau), -\xi(\tau))]$$

hold.

Then for every  $l$  the class  $K[G=l]$  either is empty or contains a curve  $C$  for which  $\mathcal{F}(C)$  assumes its least value on  $K[G=l]$ .

Sections 3 to 5 will be devoted to the proof of this theorem, and throughout these sections the hypotheses of Theorem 2 will be assumed to hold.

**3. Construction of a minimizing sequence.** If the class  $K[G=l]$  is not empty, we denote by  $\mu$  the greatest lower bound of  $\mathcal{F}(C)$  on the class  $K[G=l]$ . Also we define  $\mu_0$  to be the greatest lower bound of numbers  $m$  for which there exists a sequence  $\{C_n\}$  of curves of  $K$  having

$$\lim_{n \rightarrow \infty} \mathcal{F}(C_n) = m, \quad \lim_{n \rightarrow \infty} G(C_n) = l.$$

Since  $\mu$  is such a number  $m$ , we evidently have  $\mu_0 \leq \mu$ . If  $\{h_n\}$  is a sequence of numbers greater than  $\mu_0$  and tending to  $\mu_0$ , for each  $n$  there is a  $C_n^*$  such that

$$\mathcal{F}(C_n^*) < h_n, \quad |G(C_n^*) - l| < 1/n.$$

Hence

$$(3.1) \quad \lim_{n \rightarrow \infty} \mathcal{F}(C_n^*) = \mu_0, \quad \lim_{n \rightarrow \infty} G(C_n^*) = l.$$

By (1.1), the  $C_n^*$  have bounded lengths, which ensures the finiteness of  $\mu_0$ .

Suppose  $C_n^*$  has the Lipschitzian representation  $z = z_n^*(t)$ ,  $(0 \leq t \leq 1)$ . The set of values of  $t$  such that  $z_n^*(t)$  has a distance greater than  $1/2n$  from the boundary of  $S$  is open relative to  $[0, 1]$  and so falls into a finite or denumerable aggregate of subintervals. For only a finite number of these, say

$[\alpha_1, \beta_1], \dots, [\alpha_n, \beta_n]$ , can the corresponding arc of  $C_n^*$  have length as great as  $1/2n$ ; the others we disregard. Thus if  $t$  is not in any of the intervals  $[\alpha_i, \beta_i]$ , the point  $z_n^*(t)$  has distance less than  $n^{-1}$  from the boundary of  $S$ .

Consider now one of the intervals  $[\alpha_i, \beta_i]$ . In the corresponding arc of  $C_n^*$  we inscribe a polygon  $\Pi_{i,n}^*$  with the successive vertices

$$z_n^*(\alpha_i), z_n^*(t_1), \dots, z_n^*(t_m), z_n^*(\beta_i), \quad \alpha_i < t_1 < \dots < t_m < \beta_i.$$

If the sides of this polygon are short enough, we will have

$$(3.2) \quad \left| \mathcal{F}(\Pi_{i,n}^*) - \int_{\alpha_i}^{\beta_i} F(z_n^*, \dot{z}_n^*) dt \right| < 1/hn,$$

and

$$(3.3) \quad \left| \mathcal{G}(\Pi_{i,n}^*) - \int_{\alpha_i}^{\beta_i} G(z_n^*, \dot{z}_n^*) dt \right| < 1/hn.$$

Now let  $\Pi_{i,n}^*$  be the polygon having the same number of vertices as  $\Pi_{i,n}^*$ , joining  $z_n^*(\alpha_i)$  to  $z_n^*(\beta_i)$ , having  $\mathcal{G}(\Pi_{i,n}) = \mathcal{G}(\Pi_{i,n}^*)$ , and minimizing  $\mathcal{F}(C)$  in that class of polygons. Such a polygon exists, by Lemma 3 of IV. For this polygon we have

$$(3.4) \quad \mathcal{F}(\Pi_{i,n}) < \int_{\alpha_i}^{\beta_i} F(z_n^*, \dot{z}_n^*) dt + 1/hn,$$

$$(3.5) \quad \left| \mathcal{G}(\Pi_{i,n}) - \int_{\alpha_i}^{\beta_i} G(z_n^*, \dot{z}_n^*) dt \right| < 1/hn.$$

Now we define the curve  $C_n$  to be the curve  $C_n^*$  with the arcs corresponding to the intervals  $[\alpha_i, \beta_i]$  replaced by the respective polygons  $\Pi_{i,n}$ . Since we can consider the functions  $z_n^*$  unaltered except on the intervals  $[\alpha_i, \beta_i]$ , we obtain from (3.4) and (3.5) the relations

$$(3.6) \quad \mathcal{F}(C_n) < \mathcal{F}(C_n^*) + 1/n,$$

$$(3.7) \quad |\mathcal{G}(C_n) - \mathcal{G}(C_n^*)| < 1/n.$$

Hence

$$(3.8) \quad \limsup_{n \rightarrow \infty} \mathcal{F}(C_n) \leq \lim_{n \rightarrow \infty} \mathcal{F}(C_n^*) + 0 = \mu_0,$$

$$(3.9) \quad \lim_{n \rightarrow \infty} \mathcal{G}(C_n) = l.$$

But by the definition of  $\mu_0$  we cannot have  $\liminf \mathcal{F}(C_n) < \mu_0$ , in the presence of (3.9). So from this and (3.8) we conclude that

$$(3.10) \quad \lim_{n \rightarrow \infty} \mathcal{Y}(C_n) = \mu_0 \leq \mu.$$

By (1.1), the curves  $C_n$  are of uniformly bounded lengths; so we can select a subsequence for which  $\mathcal{L}(C_n)$  converges to a finite limit  $L$ :

$$(3.11) \quad \lim_{n \rightarrow \infty} \mathcal{L}(C_n) = L < \infty.$$

Clearly we may suppose (since we can discard a finite number of the  $C_n$ ) that we have

$$(3.12) \quad \mathcal{L}(C_n) \leq L + 1.$$

For this subsequence (3.9) and (3.10) still hold.

On each curve  $C_n$  we introduce as parameter  $t = s/\mathcal{L}(C_n)$ , where  $s$  is arc length on  $C_n$ . Then  $C_n$  has the representation

$$C_n: \quad z = z_n(t), \quad 0 \leq t \leq 1.$$

These functions  $z_n(t)$  satisfy a Lipschitz condition of constant  $\mathcal{L}(C_n)$ , which is less than or equal to  $L+1$  by (3.12). Hence

$$(3.13) \quad |\dot{z}_n(t)| \leq L + 1, \quad 0 \leq t \leq 1.$$

By Ascoli's theorem, we can select a subsequence of the  $C_n$  (we suppose it the whole sequence) such that the functions  $z_n(t)$  converge uniformly to a limit function  $z_0(t)$ :

$$(3.14) \quad \lim_{n \rightarrow \infty} z_n(t) = z_0(t) \quad \text{uniformly for } 0 \leq t \leq 1.$$

For this subsequence (3.9), (3.10), and (3.11) remain valid. The curve  $z = z_0(t)$ , ( $0 \leq t \leq 1$ ), we denote by  $C_0$ .

Next we define

$$(3.15) \quad \phi_n(t) = \int_0^t F(z_n(t), \dot{z}_n(t)) dt, \quad n = 0, 1, \dots,$$

$$(3.16) \quad \gamma_n(t) = \int_0^t G(z_n(t), \dot{z}_n(t)) dt, \quad n = 0, 1, \dots$$

These functions all satisfy the same Lipschitz condition, since the integrands are bounded. So we may select a subsequence (we again denote it by  $\{C_n\}$ ) such that  $\phi_n(t)$  and  $\gamma_n(t)$  tend uniformly to limit functions  $\phi(t)$  and  $\gamma(t)$ , respectively:

$$(3.17) \quad \lim_{n \rightarrow \infty} \phi_n(t) = \phi(t), \quad \lim_{n \rightarrow \infty} \gamma_n(t) = \gamma(t)$$

uniformly for  $0 \leq t \leq 1$ . For this subsequence (3.9), (3.10), (3.11), and (3.14) remain valid.

Equations (3.9), (3.10), (3.15), (3.16), and (3.17) imply

$$(3.18) \quad \phi(1) = \mu_0, \quad \gamma(1) = l.$$

From the manner of constructing the curves  $C_n$ , we notice that for every  $\epsilon > 0$  all the curves  $C_n$  with large  $n$  are polygons except for their arcs which are within the  $\epsilon$ -neighborhood of the boundary of  $S$ . So if  $z_0(t_0)$  is a point of  $C_0$  interior to  $S$ , there is a neighborhood  $(t_0 - \delta, t_0 + \delta)$  of  $t_0$  such that the functions  $z_n(t)$ ,  $(t_0 - \delta < t \leq t_0 + \delta)$ , represent polygonal arcs whenever  $n$  is greater than a certain  $n_0$ . (Modifications if  $t_0 = 0$  or  $1$  are obvious.)

4. **Some lemmas.** Let us first dispose of the trivial case for which we have  $L = \lim \mathcal{L}(C_n) = 0$ . This is only possible if  $z_1 = z_2$ , and it implies that  $C_0$  has length 0 and consists of the one point  $z_1$ . By (3.9) and (3.10) we have  $l = 0 = \mu_0 \leq \mu$ . By trivial computation we obtain  $G(C_0) = 0 = l$ ,  $\mathcal{I}(C_0) = 0 \leq \mu$ . Since  $C_0$  is thus in  $K[G=l]$ ,  $\mathcal{I}(C_0) \geq \mu$ ; whence  $\mathcal{I}(C_0) = \mu = \mu_0 = 0$  and  $C_0$  is the curve sought. This leaves for consideration the principal case, in which

$$(4.1) \quad L = \lim \mathcal{L}(C_n) > 0.$$

Since  $\phi$ ,  $\phi_0$ ,  $\gamma$ ,  $\gamma_0$ , and  $z_0$  are all Lipschitzian functions of  $t$ , the interval  $[0, 1]$  contains a set  $E$  of measure 1 such that for all  $t$  in  $E$  all the functions mentioned have derivatives and

$$(4.2) \quad \phi'_0(t) = F(z_0(t), z'_0(t)), \quad \gamma'_0(t) = G(z_0(t), z'_0(t)).$$

It will be supposed (without loss of generality) that neither 0 nor 1 is in  $E$ . We now begin to prove a sequence of lemmas.

LEMMA 1. *If  $t_0$  is in  $E$ , and  $a$  and  $b$  are numbers such that  $aF(z, r) + bG(z, r)$  is p.q.r. at  $z_0(t_0)$ , then*

$$a[\phi'(t_0) - \phi'_0(t_0)] + b[\gamma'(t_0) - \gamma'_0(t_0)] \geq 0.$$

Except for notation, this is merely a restatement of Lemma 5 of III.

LEMMA 2. *If  $t$  is in  $E$  and  $z(t)$  is an ordinary point, then*

$$(4.3) \quad \phi'(t_0) = \phi'_0(t_0), \quad \gamma'(t_0) = \gamma'_0(t_0).$$

By (2.2), (a),  $z_0(t_0)$  is interior to  $S$ . Therefore, as remarked at the end of §3, for all large  $n$  the arcs of  $C_n$  lying in a neighborhood of  $z_0(t_0)$  are polygonal. All the hypotheses leading up to equations (8.2) and (8.3) of IV (which together are (4.3) above) are here satisfied, except that in IV the set  $S$  was the whole space. However, the proof of (8.2) and (8.3) was purely local in nature; the only reason for taking the whole space for  $S$  was to be sure that each point of

$C_0$  should be interior to  $S$ . So the proofs in IV are applicable without change, and the lemma is established.

LEMMA 3. *If  $z_0(t_0)$  is a singular point, and  $t_0$  belongs to  $E$  but does not belong either to  $T_+(C_0)$  or to  $T_-(C_0)$ , then  $\gamma'(t_0) = \gamma'_0(t_0)$  and  $\phi'(t_0) \geq \phi'_0(t_0)$ .*

If  $G(z, r)$  were not p.q.r. at  $z_0(t_0)$ , the point  $t_0$  would be in  $T_-(C_0)$ . If it were not n.q.r. at  $z_0(t_0)$ , then  $t_0$  would be in  $T_+(C_0)$ . So  $G(z, r)$  must be both p.q.r. and n.q.r. at  $z_0(t_0)$ . That is,  $G(z_0(t_0), r)$  is linear in the variables  $r$ . By hypothesis (2.3),  $F(z, r) - \theta G(z, r)$  is p.q.r. at  $z_0(t_0)$  for some  $\theta$ . But since  $G(z, r)$  is merely linear in the  $r^i$ , this implies that  $F(z, r)$  itself is p.q.r. at  $z_0(t_0)$ . The application of Lemma 1 with  $a=1$ ,  $b=0$  yields  $\phi'(t_0) - \phi'_0(t_0) \geq 0$ . Since  $G(z, r)$  is both p.q.r. and n.q.r. at  $z_0(t_0)$ , we apply Lemma 1 with  $a=0$ ,  $b=1$  and with  $a=0$ ,  $b=-1$ . This yields two inequalities which together imply  $\gamma'(t_0) = \gamma'_0(t_0)$ , completing the proof.

LEMMA 4. *If  $t_0$  is in  $ET_+(C_0)$ , then*

$$\phi'(t_0) - \phi'_0(t_0) \geq M(z_0(t_0))[\gamma'(t_0) - \gamma'_0(t_0)].$$

By definition of  $M(z)$ , there is a sequence  $\{\theta_n\}$  of numbers tending to  $M(z_0(t_0))$  such that for each  $n$ ,  $F(z, r) - \theta_n G(z, r)$  is p.q.r. at  $z_0(t_0)$ . By lemma 1,

$$\phi'(t_0) - \phi'_0(t_0) \geq \theta_n[\gamma'(t_0) - \gamma'_0(t_0)].$$

Letting  $n \rightarrow \infty$  establishes the desired inequality.

LEMMA 5. *If  $t_0$  is in  $ET_-(C_0)$ , then*

$$\phi'(t_0) - \phi'_0(t_0) \geq m(z_0(t_0))[\gamma'(t_0) - \gamma'_0(t_0)].$$

Choose a sequence  $\{\theta_n\}$  such that  $\theta_n \rightarrow m(z_0(t_0))$  and  $F(z, r) - \theta_n G(z, r)$  is p.q.r. at  $z_0(t_0)$  for each  $n$ . The rest of the proof is a repetition of that of Lemma 4.

LEMMA 6. *If  $t_0$  is in  $E$  and  $\gamma'(t_0) = \gamma'_0(t_0)$ , then  $\phi'(t_0) \geq \phi'_0(t_0)$ .*

If  $z_0(t_0)$  is ordinary, this follows from Lemma 2. If  $z_0(t_0)$  is singular, either it is in neither  $T_+(C_0)$  nor  $T_-(C_0)$ , in which case  $\phi'(t_0) \geq \phi'_0(t_0)$  by Lemma 3, or it is in one (or both) of the sets  $T_+(C_0)$  and  $T_-(C_0)$ , in which case  $\phi'(t_0) - \phi'_0(t_0) \geq 0$  by Lemma 4 or Lemma 5.

LEMMA 7. *If  $t_0$  is in  $E$  and  $\gamma'(t_0) > \gamma'_0(t_0)$ , then  $t_0$  is in  $T_+(C_0)$ ; if  $t_0$  is in  $E$  and  $\gamma'(t_0) < \gamma'_0(t_0)$ , then  $t_0$  is in  $T_-(C_0)$ .*

The point  $z_0(t_0)$  must be singular by Lemma 2. If  $t_0$  is not in  $T_+(C_0)$ , then  $G(z, r)$  is n.q.r. at  $z_0(t_0)$ . Applying Lemma 1 with  $a=0$ ,  $b=-1$ , we obtain  $\gamma'(t_0) - \gamma'_0(t_0) \leq 0$ . Hence if  $\gamma'(t_0) - \gamma'_0(t_0) > 0$ , then  $t_0$  is in  $T_+(C_0)$ . The proof of the other statement is similar.

5. **Proof of the theorem.** We now subdivide the set  $E$  into three subsets. The set  $T_0$  will be the subset of  $E$  on which  $\gamma'(t) = \gamma'_0(t)$ ; the set  $T_1$  will be the subset of  $E$  on which  $\gamma'(t) > \gamma'_0(t)$ ; and the set  $T_2$  will be the subset of  $E$  on which  $\gamma'(t) < \gamma'_0(t)$ . These sets are clearly measurable, since  $\gamma'(t)$  and  $\gamma'_0(t)$  are measurable functions. By Lemma 7,  $T_1$  is contained in  $T_+(C_0)$  and  $T_2$  is contained in  $T_-(C_0)$ .

First we shall construct a curve  $\Gamma_1: z = \zeta_1(\tau)$ , ( $0 \leq \tau \leq \tau_1$ ), beginning and ending at a point  $z_0(t_1)$  on  $C_0$ , and such that

$$(5.1) \quad \mathcal{G}(\Gamma_1) = \int_{T_1} [\gamma'(t) - \gamma'_0(t)] dt,$$

$$(5.2) \quad \mathcal{F}(\Gamma_1) \leq \int_{T_1} [\phi'(t) - \phi'_0(t)] dt.$$

Whenever  $T_1$  is empty we can take  $\Gamma_1$  to be a degenerate curve consisting of a single point on  $C_0$ . Then (5.1) and (5.2) obviously hold. If  $T_1$  is not empty, then  $T_+(C_0)$  is also not empty. By hypothesis (c) there is a curve  $\Gamma^*$  corresponding to  $C_0$  and having the properties there specified. For  $0 \leq \tau \leq \epsilon$  we define  $\Gamma(\tau)$  to be the curve obtained by traversing  $\Gamma^*$  from  $\zeta(0)$  to  $\zeta(\tau)$  and then returning to  $\zeta(0)$ . Thus  $\Gamma(\tau)$  is defined by the equations

$$z = \zeta(t), \quad 0 \leq t \leq \tau; \quad z = \zeta(2\tau - t), \quad \tau < t \leq 2\tau.$$

This is a rectifiable continuous curve beginning and ending at  $\zeta(0)$ , which, by hypothesis (c), is a point  $z_0(t_1)$  on  $C_0$ . We calculate

$$(5.3) \quad \begin{aligned} \mathcal{G}(\Gamma(\tau)) &= \int_0^\tau G(\zeta(t), \zeta(t)) dt + \int_\tau^{2\tau} G(\zeta(2\tau - t), -\zeta(2\tau - t)) dt \\ &= \int_0^\tau [G(\zeta(t), \zeta(t)) + G(\zeta(t), -\zeta(t))] dt. \end{aligned}$$

Likewise

$$(5.4) \quad \mathcal{F}(\Gamma(\tau)) = \int_0^\tau [F(\zeta(t), \zeta(t)) + F(\zeta(t), -\zeta(t))] dt.$$

By (2.6) the integrand in (5.3) is almost everywhere positive; hence  $\mathcal{G}(\Gamma(\epsilon)) > 0$ . Let  $m$  be an integer for which

$$m \mathcal{G}(\Gamma(\epsilon)) > \int_{T_1} [\gamma'(t) - \gamma'_0(t)] dt.$$

Since  $\mathcal{G}(\Gamma(\tau))$  is a continuous function of  $\tau$ , there is a  $\tau_0$  such that

$$G(\Gamma(\tau_0)) = m^{-1} \int_{T_1} [\gamma'(t) - \gamma'_0(t)] dt.$$

We now define  $\Gamma_1$  to be the curve obtained by traversing  $\Gamma(\tau_0)$  a total of  $m$  times. Then (5.1) holds.

Recalling that  $M(C_0)$  is the greatest lower bound of  $M(z_0(t))$  for all  $t$  in  $T_+(C_0)$ , by Lemma 4 we find that whenever  $t_0$  is in  $T_1$  the inequality

$$\phi'(t_0) - \phi'_0(t_0) \geq M(C_0) [\gamma'(t_0) - \gamma'_0(t_0)]$$

holds. Hence

$$(5.5) \quad \int_{T_1} [\phi'(t) - \phi'_0(t)] dt \geq M(C_0) \int_{T_1} [\gamma'(t) - \gamma'_0(t)] dt.$$

On the other hand, by (2.7), (5.3), and (5.4) we find

$$\begin{aligned} \mathcal{J}(\Gamma_1) &= m \int_0^{\tau_0} [F(\xi(t), \dot{\xi}(t)) + F(\xi(t), -\dot{\xi}(t))] dt \\ (5.6) \quad &\leq mM(C_0) \int_0^{\tau_0} [G(\xi(t), \dot{\xi}(t)) + G(\xi(t), -\dot{\xi}(t))] dt \\ &= M(C_0) G(\Gamma_1). \end{aligned}$$

From (5.1), (5.5), and (5.6) we obtain (5.2).

Next we prove that there is a curve  $\Gamma_2$  beginning and ending at a point  $z_0(t_2)$  on  $C_0$  and such that

$$(5.7) \quad G(\Gamma_2) = \int_{T_2} [\gamma'(t) - \gamma'_0(t)] dt,$$

$$(5.8) \quad \mathcal{J}(\Gamma_2) \leq \int_{T_2} [\phi'(t) - \phi'_0(t)] dt.$$

We could prove this as we did (5.1) and (5.2). But it is much simpler to observe that if we replace  $G(z, r)$  by  $-G(z, r)$ , then  $\gamma(t)$  is replaced by  $-\gamma(t)$  and  $M(C_0)$  by  $-m(C_0)$ , while hypotheses (c) and (d) are interchanged. Then (5.7) and (5.8) are merely (5.1) and (5.2) as rewritten for  $F$  and  $-G$  in place of  $F$  and  $G$ .

We can now define the minimizing curve  $\bar{C}$ . Suppose to be specific that  $t_1 \leq t_2$ . We obtain  $\bar{C}$  by traversing  $C_0$  from  $t=0$  to  $t=t_1$ , traversing  $\Gamma_1$ , continuing along  $C_0$  from  $t=t_1$  to  $t=t_2$ , traversing  $\Gamma_2$ , then proceeding along  $C_0$  from  $t=t_2$  to  $t=1$ . We therefore have, by (3.16), (5.1), (5.7), the definition of  $T_0$ , and (3.18),

$$\begin{aligned}
 \mathcal{G}(\bar{C}) &= \mathcal{G}(C_0) + \mathcal{G}(\Gamma_1) + \mathcal{G}(\Gamma_2) \\
 &= \int_{T_0} + \int_{T_1} + \int_{T_2} \gamma'_0(t) dt + \int_{T_1} [\gamma'(t) - \gamma'_0(t)] dt \\
 (5.9) \quad &+ \int_{T_2} [\gamma'(t) - \gamma'_0(t)] dt \\
 &= \int_E \gamma'(t) dt = \int_0^1 \dot{\gamma}(t) dt = \gamma(1) = l.
 \end{aligned}$$

Similarly, using (3.15), (5.2), (5.8), the definition of  $T_0$ , Lemma 6, and (3.18), we obtain

$$\begin{aligned}
 \mathcal{J}(\bar{C}) &= \mathcal{J}(C_0) + \mathcal{J}(\Gamma_1) + \mathcal{J}(\Gamma_2) \\
 &\leq \int_{T_0} + \int_{T_1} + \int_{T_2} \phi'_0(t) dt + \int_{T_1} [\phi'(t) - \phi'_0(t)] dt \\
 (5.10) \quad &+ \int_{T_2} [\phi'(t) - \phi'_0(t)] dt \\
 &\leq \int_E \phi'(t) dt = \int_0^1 \dot{\phi}(t) dt = \phi(1) = \mu_0.
 \end{aligned}$$

But by (5.9) the curve  $\bar{C}$  is in  $K[G=l]$ ; so  $\mathcal{J}(\bar{C}) \geq \mu \geq \mu_0$ . This, with (5.10), implies

$$(5.11) \quad \mathcal{J}(\bar{C}) = \mu_0 = \mu,$$

and the proof of the theorem is complete.

Incidentally we have proved that under the hypotheses of Theorem 2 the equation  $\mu_0 = \mu$  holds. It follows with little difficulty that the value of  $\mu$ , considered as a function of  $l$ , is lower semicontinuous.

**6. Corollaries and examples.** Let us define  $T_+^*(C)$  and  $T_-^*(C)$  by deleting the words " $z(t)$  is a singular point and" in (2.3) and (2.4), and let  $M^*(C)$  and  $m^*(C)$  be the numbers defined by replacing  $T_+(C)$ ,  $T_-(C)$  by  $T_+^*(C)$ ,  $T_-^*(C)$ , respectively, in (2.5). Then if  $m(C)$  is defined, so is  $m^*(C)$ ; and  $m^*(C) \geq m(C)$ , for  $T_-^*(C)$  contains  $T_-(C)$ . Likewise, if  $M(C)$  is defined, so is  $M^*(C)$ ; and  $M^*(C) \leq M(C)$ . The following corollary is then immediately evident:

**COROLLARY 1.** *If the hypotheses of Theorem 2 hold with  $m^*(C)$ ,  $M^*(C)$  in place of  $m(C)$  and  $M(C)$ , respectively, then the class  $K[G=l]$  either is empty or contains a curve for which  $\mathcal{J}(C)$  assumes its least value on  $K[G=l]$ .*

For if the hypotheses of Corollary 1 are satisfied, so are the hypotheses of Theorem 2.

In my dissertation<sup>†</sup> I established an existence theorem which overlaps considerably with Corollary 1 but neither contains it nor is contained in it. Nor does Corollary 1 cover the five existence theorems for isoperimetric problems given by Tonelli;<sup>‡</sup> for Tonelli allows his class  $K$  to be a "complete class of total ramification," where our class  $K$  consists of the family of curves in  $S$  joining two fixed points. § In all other respects, however, Corollary 1 contains Tonelli's theorems. Take for example Tonelli's Theorem 3 (p. 473) whose generalization to  $q$  dimensions is as follows:

*Let  $S$  be bounded and closed, and let  $K$  be a complete class of curves of total ramification lying in  $S$ . Let  $F(z, r)$  be p.q.r. on  $S$ , and let  $G(z, r) = g(z)G(z, r) + a_\alpha(z)r^\alpha$ , where  $g(z)$  is nonnegative [nonpositive] on  $S$ , and through each point  $z_1$  of  $S$  there passes an arc  $\Gamma^*$  on which  $g(z) > g(z_1)$  [ $g(z) < g(z_1)$ ], provided that any continuous curve at all passes through  $z_1$ . Then  $K[G=l]$  either is empty or contains a curve for which  $\mathcal{F}(C)$  assumes its least value on  $K[G=l]$ .*

We disregard the statements in brackets; they interchange with the unbracketed statements if  $G$  is replaced by  $-G$ , which replacement does not affect the hypotheses of Theorem 2 or Corollary 1. The set  $T_+^*(C)$  consists of all  $t$  at which  $g(z(t)) > 0$  and  $F(z, r)$  is not linear in the  $r^i$ . The set  $T_-^*(C)$  is empty; so hypothesis (d) is satisfied. For each  $t$  in  $T_+^*(C)$  the function

$$F(z, r) - \theta G(z, r) = F(z, r)[1 - \theta g(z)] - \theta a_\alpha r^\alpha$$

is p.q.r. for all  $\theta \leq 1/g(z)$ . Hence  $M(z) = 1/g(z)$ . If  $T_+^*(C)$  is not empty, then the greatest lower bound of  $M(z(t))$  for  $t$  in  $T_+^*(C)$  is at least g.l.b.  $[1/g(z(t))]$ . That is,  $M^*(C) \geq 1/\max g(z(t))$ . Let  $z_1 = z(t_1)$  be a point at which  $g(z(t))$  assumes its maximum (greater than zero), and let  $\Gamma^*$  be the curve along which  $g(z) \geq g(z_1)$ . Then along  $\Gamma^*$  the conditions (2.6) and (2.7) are satisfied. Hypotheses (1.1) and (2.2) obviously hold. So except for the added generality of the class  $K$ , this theorem is contained in Corollary 1.

As an example covered by Theorem 2 but not by Corollary 1 or any other theorems cited, let us consider

$$\begin{aligned} F(x, y, x', y') &= -(x - y)^2(x'^2 + 8y'^2)^{1/2}, \\ G(x, y, x', y') &= (x'^2 + y'^2)^{1/2}, \end{aligned}$$

<sup>†</sup> *Semi-continuity in the calculus of variations, and absolute minima for isoperimetric problems*, published in *Contributions to the Calculus of Variations*, 1930, Chicago, 1931, pp. 199-243, in particular p. 220.

<sup>‡</sup> L. Tonelli, *Fondamenti di Calcolo delle Variazioni*, vol. 2, pp. 466-482.

§ It would, however, be quite easy to extend Corollary 1 to cover such classes  $K$  of curves. The only reason for not considering them in the first place was that the discussion of ordinary points required comparison curves other than those obtained by adding a spur like  $\Gamma_1$  or  $\Gamma_2$  to a given curve. In Corollary 1 the characteristic properties of ordinary points are ignored; so this need disappears.

where the range  $S$  of  $(x, y)$  is the whole plane. Hypothesis (d) of Corollary 1 is not satisfied; inequality (2.7) cannot be satisfied unless  $C$  lies entirely along the line  $y=x$ . However, every point  $(x, y)$  not on the line  $y=x$  is an ordinary point. For, first,  $G$  is regular. Second, the matrix  $\Delta(x, y; p_1, p_2; q_1, q_2)$  has the form

$$\begin{pmatrix} -(y-x)\{p_1(p_1^2+8p_2^2)^{-1/2} & -8(y-x)\{p_2(p_1^2+8p_2^2)^{-1/2} \\ & -q_1(q_1^2+8q_2^2)^{-1/2}\} & -q_2(q_1^2+8q_2^2)^{-1/2}\} \\ p_1-q_1 & p_2-q_2 \end{pmatrix}$$

if we assume (as we may) that  $p_1^2+p_2^2=q_1^2+q_2^2=1$ . If a vector  $(p_1, p_2)$  is given, the vector  $(q_1, q_2)=(p_1, -p_2)$  is an approach set containing  $(p_1, p_2)$ , and  $(-p_1, p_2)$  is in another approach set containing  $(p_1, p_2)$ . It is possible (though not very easy) to show that no approach set contains any other unit vectors than these. Computing  $\Omega_H$  we see that it is not zero for any of the sets except in the trivial case in which the two formally different unit vectors of the approach set coincide ( $p_1=0$  or  $p_2=0$ ). Hence every point of  $S$  with  $y \neq x$  is an ordinary point.

The singular points of  $S$  are thus the points  $(x, x)$ . For these the function  $F(x, y, x', y') - \theta G(x, y, x', y')$  reduces to  $-\theta(x'^2+y'^2)^{1/2}$ , which is positive quasi-regular if and only if  $\theta \leq 0$ . So  $m(x, x) = -\infty$  and  $M(x, x) = 0$ . The set  $T_-(C)$  is always empty, since  $G(z, r)$  is positive regular. If  $T_+(C)$  is not empty, then for every  $t$  in  $T_+(C)$  we have  $x(t) = y(t)$  and  $M(x(t), y(t)) = 0$ . Therefore  $M(C)$  is 0 whenever it is defined; that is, whenever  $C$  intersects the line  $y=x$ . Thus if  $C$  does not intersect the line, then  $T_+(C)$  is empty; and if  $C$  intersects the line at a point  $(x_0, x_0)$ , we can take  $\Gamma^*$  to be a segment of  $y=x$  beginning at  $(x_0, x_0)$ . Hypothesis (d) therefore is satisfied.

If we use the same  $G$ , but take

$$F = e^{(y+x)^2}(x'^2+8y'^2)^{1/2}$$

and let  $S$  be the whole  $(x, y)$ -plane, we find similarly that there are no singular points.

**7. A generalization.** There are several ways of strengthening Theorem 2 without great difficulty. An obvious one is as follows: If  $z_1, z_2$ , and  $l$  are given, under hypothesis (1.1) there is only a bounded subset of  $S$  which can contain points of curves  $C$  of  $K[G=l]$  with  $\mathcal{F}(C) \leq \mu + \epsilon$  for any given  $\epsilon > 0$ . Let  $S$ , be this subset. We need then assume only that hypotheses (c) and (d) of Theorem 2 hold on the closure of  $S$ .

A less trivial generalization is obtained by redefining  $M(z)$  and  $m(z)$  at singular points  $z$  which are interior to  $S$ . Let  $z$  be such a singular point, and

let  $A$  be an approach set at  $z$ . Consider the aggregate of numbers  $\theta$  for which

$$(7.1) \quad a_\beta \mathcal{E}_{F-\theta G}(z, a_\alpha p_\alpha, p_\beta) \geq 0$$

for all finite collections  $p_1, \dots, p_n$  of vectors of  $A$  and all sets  $a_1, \dots, a_n$  of nonnegative numbers such that  $|a_\alpha p_\alpha| \neq 0$ . We make the following definitions:

(7.2)  $M_1(z, A)$  is the least upper bound of all numbers  $\theta$  such that (7.1) holds, and  $m_1(z, A)$  is their greatest lower bound.

(7.3)  $M_1(z)$  is the greatest lower bound of  $M_1(z, A)$  for all approach sets  $A$  at  $z$ , and  $m_1(z)$  is the least upper bound of  $m_1(z, A)$  for all approach sets  $A$  at  $z$ .

Under hypothesis (2.2) such numbers  $\theta$  exist. For if  $\theta$  serves in (2.2), then

$$\mathcal{E}_{F-\theta G}(z, p, r) \geq 0$$

for all  $p \neq 0$  and all  $r$ , and (7.1) follows at once. This argument shows moreover that every  $\theta$  which serves in (2.2) serves in (7.1), no matter which approach set  $A$  we use. Hence if  $A$  is any approach set at  $z$ ,  $M_1(z, A) \geq M(z)$  and  $m_1(z, A) \leq m(z)$ ; so by (7.3)

$$(7.4) \quad M_1(z) \geq M(z), \quad m_1(z) \leq m(z).$$

Our theorem is given as follows:

**THEOREM 3.** *At all singular points  $z$  interior to  $S$  let  $M(z)$ ,  $m(z)$  be redefined to mean  $M_1(z)$ ,  $m_1(z)$ , respectively. Then with this new meaning of  $m(z)$  and  $M(z)$  Theorem 2 remains valid.*

The numbers  $M(z)$ ,  $m(z)$  entered the proof of Theorem 2 by way of Lemmas 4 and 5. Therefore we need only establish Lemmas 4 and 5 with  $M_1$ ,  $m_1$  in place of  $M$ ,  $m$ , respectively. Suppose then that  $z_0(t_0)$  is a singular point interior to  $S$ . By Lemma 5 of IV there is an approach set  $A$  and a subsequence  $\{z_m(t)\}$  with the properties there specified. (We disregard case (i) of that Lemma, for then the proof that  $\phi'(t_0) = \phi'_0(t_0)$  and  $\gamma'(t_0) = \gamma'_0(t_0)$  goes through as before.) By the definition of the  $\mathcal{E}$ -function, inequality (7.1) can be written

$$(7.5) \quad a_\alpha [F(z_0, p_\alpha) - \theta G_\alpha(z_0, p_\alpha)] \geq F(z_0, a_\alpha p_\alpha) - \theta G(z_0, a_\alpha p_\alpha),$$

where we have written  $z_0$  for  $z_0(t_0)$ . Let  $R$  be the (convex) set consisting of all nonzero vectors  $r$  which can be written in the form  $a_1 p_1 + \dots + a_n p_n$ , where each  $p_i$  is in  $A$  and  $a_i \geq 0$ . Each  $r$  in  $R$  can be written in one or many ways as a sum  $a_\alpha p^\alpha$ . The lower bound of the left member of (7.5), for all such ways of writing  $r$ , is known to be a convex function  $\bar{H}(z_0, r)$  on  $R$ . By (7.5),

$$\bar{H}(z_0, r) \geq F(z_0, r) - \theta G(z_0, r), \quad r \text{ in } R.$$

From this (and the differentiability of  $F$  and  $G$ ) there is, for each  $r_0$  in  $R$ , a linear function  $l_\alpha r^\alpha$  such that

$$(7.6) \quad F(z_0, r_0) - \theta G(z_0, r_0) \leq l_\alpha r_0^\alpha,$$

$$(7.7) \quad \overline{H}(z_0, r) \geq l_\alpha r^\alpha \text{ for all } r \text{ in } R.$$

In particular,  $F(z_0, p) - \theta G(z_0, p) \geq \overline{H}(z_0, p)$  for all  $p$  in  $A$ , as we find by taking  $n=1$ ,  $p_1=p$ ,  $a_1=1$ , in (7.5). So by (7.7)

$$(7.8) \quad F(z_0, p) - \theta G(z_0, p) \geq l_\alpha p^\alpha, \quad p \text{ in } A.$$

Let us denote the closed  $\gamma$ -neighborhoods of  $z_0, A, R$  by  $(z_0)_\gamma, (A)_\gamma, (R)_\gamma$ , respectively. For each  $\gamma > 0$ , if  $m$  is large and  $\delta$  small, the point  $z_m(t)$  is  $(z_0)_\gamma$  and  $z_m'(t)$  is in  $(A)_\gamma$  for  $t_0 - \delta \leq t \leq t_0 + \delta$ . A fortiori,  $z_m'(t)$  is in the closed convex set  $(R)_\gamma$ . By Jensen's inequality, if  $t_0 - \delta \leq t < t_0 + \delta$ , then

$$\frac{1}{h} \int_t^{t+h} \dot{z}_m(t) dt \equiv [z_m(t+h) - z_m(t)]/h$$

is in  $(R)_\gamma$ . Let  $m \rightarrow \infty$ ; the vector  $[z_0(t+h) - z_0(t)]/h$  is in  $(R)_\gamma$ . Let  $h \rightarrow 0$ ; the vector  $z_0'(t)$ , if defined, is in  $(R)_\gamma$ . By use of fairly obvious estimates, we find by (7.6) and (7.8) that for all sufficiently small positive numbers  $\gamma$

$$(7.9) \quad F(z_0(t), z_0'(t)) - \theta G(z_0(t), z_0'(t)) \leq l_\alpha z_0^{\alpha'}(t) + \epsilon,$$

$$(7.10) \quad F(z_m(t), z_m'(t)) - \theta G(z_m(t), z_m'(t)) \geq l_\alpha z_m^{\alpha'}(t) - \epsilon,$$

$t_0 - \delta \leq t < t_0 + \delta, m \text{ large.}$

Integrating from  $t_0$  to  $t_0 + h$  yields

$$\begin{aligned} \phi_0(t_0 + h) - \phi_0(t_0) - \theta[\gamma_0(t_0 + h) - \gamma_0(t_0)] &\leq l_\alpha [z_0^{\alpha'}(t_0 + h) - z_0^{\alpha'}(t_0)] + \epsilon h, \\ \phi_m(t_0 + h) - \phi_m(t_0) - \theta[\gamma_m(t_0 + h) - \gamma_m(t_0)] &\geq l_\alpha [z_m^{\alpha'}(t_0 + h) - z_m^{\alpha'}(t_0)] - \epsilon h. \end{aligned}$$

If we let  $m \rightarrow \infty$ , divide by  $h$ , and let  $h \rightarrow 0$ , we get

$$\phi_0'(t_0) - \theta \gamma_0'(t_0) \leq l_\alpha z_0^{\alpha'}(t_0) + \epsilon, \quad \phi'(t_0) - \theta \gamma'(t_0) \geq l_\alpha z_0^{\alpha'}(t_0) - \epsilon.$$

Since  $\epsilon$  is arbitrary,

$$\phi'(t_0) - \phi_0'(t_0) \geq \theta[\gamma'(t_0) - \gamma_0'(t_0)].$$

If we let  $\theta$  run through a sequence of values approaching  $M_1(z_0(t_0), A)$ , we find

$$\phi'(t_0) - \phi_0'(t_0) \geq M_1(z_0(t_0), A)[\gamma'(t_0) - \gamma_0'(t_0)].$$

This holds for all  $M_1(z_0(t_0), A)$ ; so it holds for their greatest lower bound  $M_1(z_0(t_0))$ . The generalization of Lemma 4 is therefore established. Lemma 5 can be discussed similarly; or we can obtain the result from the proof above by replacing  $G$  by  $-G$ .

From the definitions it is evident that any alteration in the definitions of  $M_1(z, A)$ ,  $m_1(z, A)$ ,  $M_1(z)$ , and  $m_1(z)$  which enables us to discard vectors  $p$  from approach sets  $A$  at  $z$  or enables us to disregard entire approach sets  $A$  either leaves these numbers unchanged or improves them; that is, if  $M_1(z, A)$  and  $M_1(z)$  are altered by the change of definition they are increased, and if  $m_1(z, A)$  and  $m_1(z)$  are altered they are decreased. In §8 we shall establish a criterion which will permit us to ignore certain types of approach sets. Here we establish two simpler criteria. Suppose that  $z$  is a singular point interior to  $S$ , at which  $G(z, r)$  is quasi-regular normal. If  $A$  is an approach set containing only a finite number of unit vectors  $p_1, \dots, p_k$ , and if these can be so ordered that  $\Omega_H(z, p_i, p_j) < 0$  if  $i < j$ , then  $A$  can be disregarded in defining  $m_1(z)$  and  $M_1(z)$ . For if, in Lemma 5 of IV, the set  $A$  can be so chosen as to have these properties, all the proof leading up to equations (8.2) and (8.3) of IV remains valid without change, and we obtain  $\gamma'(t_0) = \gamma'_0(t)$  and  $\phi'(t_0) = \phi'_0(t_0)$ .

Retaining the assumption concerning  $G(z, r)$ , let us suppose that  $A$  is an approach set containing a finite number of unit vectors. If there is a unit vector  $p_1$  in  $A$  such that  $\Omega_H(z, p_1, p) < 0$  for all unit vectors  $p \neq p_1$  in  $A$ , we say that  $p_1$  is the first vector in  $A$ . If there is also a unit vector  $p_2$  in  $A$  such that  $\Omega_H(z, p_2, p) < 0$  for all unit vectors  $p$  in  $A$  except  $p_1$  and  $p_2$ , then  $p_2$  is the second vector in  $A$ ; and so on. If there is a  $p_l$  in  $A$  such that  $\Omega_H(z, p, p_l) < 0$  for all unit vectors  $p \neq p_l$  in  $A$ , we say that  $p_l$  is the last vector in  $A$ ; and so on for  $p_{l-1}, p_{l-2}, \dots$ . Unless  $z$  is an ordinary point, there may remain some unit vectors not thus classified. These and their multiples we call the non-ordered nucleus of  $A$ . In defining  $M_1(z, A)$  and  $m_1(z, A)$  we can discard from  $A$  all vectors not belonging to the non-ordered nucleus. The details of proof I shall omit.

8.  **$\mathcal{E}$ -Admissibility.** In §8 of III we introduced a concept called  $\mathcal{E}$ -admissibility, and showed that we could restrict our attention to those approach sets which were  $\mathcal{E}$ -admissible. The set  $A$  was  $\mathcal{E}$ -admissible if  $\mathcal{E}(z, p_0, p) \geq 0$  for all  $p_0$  in  $A$  and all  $p$ . If we wish to define an analogous notion for isoperimetric problems, we must be guided by the way in which the Weierstrass condition is stated for those problems. The Weierstrass condition is to the effect that  $\mathcal{E}_H(z_0(t), z'_0(t), p) \geq 0$  if  $z = z_0(t)$  is the minimizing curve and  $H = F - \lambda G$ . This suggests the following definition:

(8.1) Let  $A$  be an approach set at  $z$ , and let  $H(z, r)$  be the function  $F(z, r) - \lambda(z, A)G(z, r)$ . Then the set  $A$  is  $\mathcal{E}$ -admissible if

$$\mathcal{E}_H(z, p_0, p) \geq 0$$

for all  $p_0$  in  $A$  and all  $p$ .

In complete analogy with III, we can prove the following theorem:

**THEOREM 4.** *If in the definition of ordinary point we replace the words (in (2.1), (c)) "every approach set" by "every  $\mathcal{E}$ -admissible approach set," Theorems 2 and 3 remain valid.*

I have been unable to find any proof of this theorem which is not extremely long and involved. Therefore I shall here content myself with a sketch of a proof; the reader will probably be able to furnish the omitted details if he is interested.

Let  $\mu_0$  be (as before) the least number which is the limit of  $\mathcal{V}(\Pi_n)$  for a sequence of polygons  $\Pi_n$  of  $K$  such that  $G(\Pi_n) \rightarrow l$ . We may suppose that  $\Pi_n$  has the usual minimizing property with respect to curves having not more vertices than  $\Pi_n$  has. We thus come to Lemma 2 and must establish that lemma. Suppose the contrary, that one of the equations (4.3) fails. In particular, we suppose that the second one fails. In the proof of Lemma 1 an approach set  $A$  entered, via Lemma 5 of IV. If this approach set is  $\mathcal{E}$ -admissible, the whole argument leading to equations (4.3) is valid without alteration, and (4.3) holds. This is a contradiction. It remains to consider the possibility that  $A$  is not  $\mathcal{E}$ -admissible and show that this leads to a contradiction.

If  $A$  is not  $\mathcal{E}$ -admissible, there is a vector  $p_1$  such that  $\mathcal{E}_H(z, p_0, p_1) < 0$  for some (hence for all)  $p_0$  in  $A$ . Choose a small interval  $[t - \delta, t_0 + \delta]$ . We treat the two subintervals  $[t_0 - \delta, t_0]$  and  $[t_0, t_0 + \delta]$  differently.

If, as usual, we write  $H(z, r) = F(z, r) - \lambda(z_0(t_0), A)G(z, r)$ , we may assume  $H(z_0(t_0), p) = 0$  for all  $p$  in  $A$ ; for this may be brought about by adding the linear function  $-H_a(z_0(t_0), p_0)r^a$  to  $H(z, r)$ , where  $p_0$  is in  $A$ . Let  $\zeta_m(t, 0)$  be the linear function for which  $\zeta_m(t_0, 0) = z_m(t_0)$  and  $\zeta_m(t_0 + \delta, 0) = z_m(t_0 + \delta)$ , and define

$$\zeta_m(t, \sigma) = \zeta_m(t, 0) + \sigma[z_m(t) - \zeta_m(t, 0)], \quad t_0 \leq t \leq t_0 + \delta.$$

Thus  $\zeta_m(t, 1) \equiv z_m(t)$ . To simplify the situation we shall ignore the dependence of  $F(z, r)$  and  $G(z, r)$  on  $z$ . It is easy to verify that all the integrals over  $[t_0 - \delta, t_0]$  and  $[t_0, t_0 + \delta]$  are thereby changed by at most  $\theta(m, \delta)\delta$ , where  $\theta(m, \delta)$  tends to zero as  $m \rightarrow \infty$  and  $\delta \rightarrow 0$ . Accordingly, we write  $G(r)$  for  $G(z, r)$ , and so on.

The integral

$$I_m(\sigma) = \int_{t_0}^{t_0+\delta} G(\zeta_m(t, \sigma))dt = \int_{t_0}^{t_0+\delta} G(\zeta_m(t, 0) + \sigma[z_m(t) - \zeta_m(t, 0)])dt$$

is a convex function of  $\sigma$ , since  $G(r)$  is convex. If we write  $\gamma'(t_0) - \gamma'_0(t_0) = 3\kappa$ , then for all large  $m$  and small  $\delta$  we have

$$[\gamma_m(t_0 + \delta) - \gamma_m(t_0)] - [\gamma_0(t_0 + \delta) - \gamma_0(t_0)] > 2\kappa\delta.$$

Since  $G(r)$  is independent of  $z$  and is positive quasi-regular, the line segment  $z = \xi_m(t, 0)$  furnishes an absolute minimum for  $G(C)$  in the class of curves joining its ends. Hence

$$\lim_{m \rightarrow \infty} \int_{t_0}^{t_0+\delta} G(\xi_m) dt \leq \int_{t_0}^{t_0+\delta} G(\xi_0) dt.$$

With the preceding inequality, this shows that for large  $m$  and small  $\delta$

$$(8.2) \quad I_m(1) = \gamma_m(t_0 + \delta) - \gamma_m(t_0) \geq \int_{t_0}^{t_0+\delta} G(\xi_m) dt + \kappa\delta = I_m(0) + \kappa\delta.$$

From the convexity of  $I_m(\sigma)$  we find that

$$(8.3) \quad I_m(1 - \sigma) - I_m(1) < -\sigma\kappa\delta \quad \text{for } 0 \leq \sigma \leq 1.$$

Now we consider the interval  $(t_0 - \delta, t_0)$ . On this we use the construction of §8 of III. We thereby replace the arc  $z = z_m(t)$ ,  $(t_0 - \delta \leq t \leq t_0)$ , of  $\Pi_m$  by an arc  $z = z_m(t, \epsilon)$  with the same ends and having

$$(8.4) \quad \int_{t_0-\delta}^{t_0} H(\dot{z}_m(t, \epsilon)) dt - \int_{t_0-\delta}^{t_0} H(\dot{z}_m(t)) dt < -2\gamma\delta\epsilon, \quad \gamma > 0,$$

for all  $m$ . Because of the convexity of  $G(r)$  the integral of  $G$  is increased, but it is easy to estimate that the increase is less than  $K\delta\epsilon$ ,  $K$  a constant.

Choose  $\epsilon$  small enough so that  $K\delta\epsilon < \kappa\delta$ . Then by (8.3) there is a  $\sigma$  such that  $0 \leq \sigma \leq 1$  and

$$(8.5) \quad I_m(1 - \sigma) - I_m(1) = \int_{t_0-\delta}^{t_0} G(z'_m(t)) dt - \int_{t_0-\delta}^{t_0} G(z'_m(t, \epsilon)) dt.$$

That is, if we let  $\bar{z}_m(t)$  be  $z_m(t, \epsilon)$  on  $[t_0 - \delta, t_0]$  and  $\bar{z}_m(t, \sigma)$  on  $[t_0, t_0 + \delta]$ , then

$$(8.6) \quad \int_{t_0-\delta}^{t_0+\delta} G(\bar{z}'_m) dt = \int_{t_0-\delta}^{t_0+\delta} G(z'_m) dt.$$

The right member of equation (8.5) has a value between  $-K\delta\epsilon$  and zero; so by inequality (8.3) we conclude that

$$(8.7) \quad 0 < \sigma < K\epsilon/\kappa.$$

By reducing  $\epsilon$  if necessary, we can ensure that  $\sigma$  is less than  $1/12$ .

Since  $A$  is an approach set on which  $H$  vanishes identically, all first-order partial derivatives of  $H$  also vanish on  $A$ . Therefore there is a positive number  $\lambda$  such that

$$(8.8) \quad H_\alpha(r)H_\alpha(r) < [\gamma\kappa/3KL]^2$$

if  $L/2 \leq |r| \leq 3L/2$  and  $r$  is in the  $2\lambda$ -neighborhood of  $A$ . If  $m$  is large, the inequality

$$(8.9) \quad 3L/4 < |z'_m(t)| < 5L/4$$

holds; and moreover  $z'_m(t)$  is in the  $\lambda$ -neighborhood of  $A$  if  $|t-t_0| < \delta$ , provided that  $\delta$  is small enough. By definition of  $\zeta_m$  we find that

$$(8.10) \quad \begin{aligned} |\zeta'_m(t, 1-\sigma) - z'_m(t)| &= \sigma |z'_m(t) - \zeta'_m(t, 0)| \\ &< \sigma 3L < L/4. \end{aligned}$$

By using the theorem of the mean, together with (8.7) and (8.8), whose use is permitted by (8.9) and (8.10), we obtain

$$(8.11) \quad |H(\zeta'_m(t, 1-\sigma)) - H(z'_m(t))| < \sigma \gamma \kappa / K < \gamma \epsilon.$$

Recalling the definition of  $\bar{z}_m(t)$ , inequalities (8.4) and (8.11) (integrated from  $t_0$  to  $t_0+\delta$ ), we obtain

$$(8.12) \quad \begin{aligned} \int_{t_0-\delta}^{t_0+\delta} F(\bar{z}'_m) dt - \int_{t_0-\delta}^{t_0+\delta} F(z'_m) dt &= \int_{t_0-\delta}^{t_0+\delta} H(\bar{z}'_m) dt - \int_{t_0-\delta}^{t_0+\delta} H(z'_m) dt \\ &< -\gamma \delta \epsilon. \end{aligned}$$

Extending  $\bar{z}_m(t)$  by setting it equal to  $z_m(t)$  for  $0 \leq t < t_0 - \delta$  and  $t_0 + \delta < t \leq 1$ , we obtain a polygon  $\bar{\Pi}_m$  such that

$$G(\bar{\Pi}_m) = G(\Pi_m), \quad \mathcal{Y}(\bar{\Pi}_m) < \mathcal{Y}(\Pi_m) - \gamma \delta \epsilon.$$

But then  $\limsup \mathcal{Y}(\Pi_m) \leq \mu_0 - \gamma \delta \epsilon$ , contrary to the definition of  $\mu_0$ , and the desired contradiction has been reached.

We thus see that  $\gamma'(t_0) = \gamma'_0(t_0)$  for almost all  $t_0$  such that  $z_0(t_0)$  is interior to  $S$ . If  $t_1 \leq t \leq t_2$  defines an interior arc of  $z = z_0(t)$ , this proves that

$$\lim_{n \rightarrow \infty} \int_{t_1}^{t_2} G(z_n, \dot{z}_n) dt = \int_{t_1}^{t_2} G(z_0, \dot{z}_0) dt.$$

But  $G$  is quasi-regular normal; so by a known theorem this implies

$$\lim_{n \rightarrow \infty} \int_{t_1}^{t_2} F(z_n, \dot{z}_n) dt = \int_{t_1}^{t_2} F(z_0, \dot{z}_0) dt;$$

that is,  $\phi(t_2) - \phi(t_1) = \phi_0(t_2) - \phi_0(t_1)$ . Hence  $\phi'(t_0) = \phi'_0(t_0)$  for all points  $z_0(t_0)$  interior to  $S$ , and equations (4.3) are established.

Besides the added generality, Theorem 4 offers another advantage. The search for  $\mathcal{E}$ -admissible approach sets may be easier than the determination of all approach sets. For one thing, if  $A$  is an approach set at  $z$ , and there is a  $z'$  such that

$$(8.13) \quad H(z, z'; \lambda(z, A)) + H(z, -z'; \lambda(z, A)) < 0,$$

then  $A$  is not  $\mathcal{E}$ -admissible, as we see if we rewrite (8.13) in the form

$$\mathcal{E}_H(z, p, z') + \mathcal{E}_H(z, p, -z') < 0.$$

If  $F$  and  $G$  belong to the important special class of integrands such that

$$F(z, -z') = F(z, z'), \quad G(z, -z') = G(z, z'),$$

then in order that the approach set  $A$  at  $z$  be  $\mathcal{E}$ -admissible it is necessary that  $H(z, z'; \lambda(z, A))$  be nonnegative. The example of §6 is of this type. More generally, let

$$F = \phi(x, y)(x'^2 + a^2y'^2)^{1/2}, \quad G = \psi(x, y)(x'^2 + y'^2)^{1/2},$$

where  $a > 1$  and  $\psi > 0$ . Here

$$H = \phi(x'^2 + a^2y'^2)^{1/2} - \lambda\psi(x'^2 + y'^2)^{1/2},$$

and in order that this be nonnegative we must have

$$(8.14) \quad \lambda \leq a\phi/\psi.$$

Suppose to be specific that  $\phi \leq 0$ . If the equality holds in (8.14), then  $H(x, y, 0, y'; \lambda) = 0$  for all  $y'$ , and  $(0, 1)$  and  $(0, -1)$  are in an  $\mathcal{E}$ -admissible approach set. No other unit vectors are in this set unless  $\phi = 0$ . If  $\lambda < a\phi/\psi$  then  $H$  is positive. The graph of  $H = 1$  is either convex (if  $-\lambda$  is large) or dumb-bell shaped, with its narrowest section along the  $x'$ -axis. It is then geometrically evident that the only  $\mathcal{E}$ -admissible approach sets are those containing only two unit vectors,  $(p, q)$  and  $(p, -q)$ . This applies, in particular, to the example of §6.

Again, let  $F = -e^y(x'^2 + 4y'^2)^{1/2}$ ,  $G = (x'^2 + y'^2)^{1/2}$ . As we have just seen, the only  $\mathcal{E}$ -admissible approach sets contain at most two unit vectors,  $(p, q)$  and  $(p, -q)$ . Suppose  $q > 0$ ; then  $\Omega_H(x, y, p, q, p, -q) = -e^y\{2q(p^2 + 4q^2)^{1/2}\}$ . So (2.1), (c) holds for the  $\mathcal{E}$ -admissible approach sets  $A$ , and Theorem 4 applies. But Theorems 1 and 2 do not apply. For if  $(p, q)$  is a unit vector, then  $(-p, q)$  is in an approach set with  $(p, q)$ , and  $\Omega_H(x, y, p, q, -p, q) = 0$ . Hence (2.1) is not satisfied for all approach sets  $A$ , and no point is ordinary as defined in (2.1).

UNIVERSITY OF VIRGINIA,  
CHARLOTTESVILLE, VA.

# ON A CALCULUS OF OPERATORS IN REFLEXIVE VECTOR SPACES\*

BY

EDGAR R. LORCH

## I. INTRODUCTION

Linear transformation theory in general vector spaces is not nearly as extensive as it is for that special space, Hilbert space. In Hilbert space large and important classes of transformations, the self-adjoint and unitary transformations, may be studied exhaustively because these transformations are susceptible of a spectral resolution. In turn, the spectral theory leads to an elegant calculus of these transformations or operators which asserts the existence of a ring homomorphism between a class of functions and a class of permutable operators centered about a given operator. These developments are possible because Hilbert space is self-adjoint.

Other vector spaces have not to the present yielded such rich results. The operators of no important class have been found to be completely resolvable. Indeed, a theory of projections which is the first step toward a spectral development has not to our knowledge been given, although the matter has received attention before this. In his investigations in the problem of complementary manifolds in the spaces  $L_p$  and  $l_p$ , F. J. Murray introduces the notion of projection at an early stage.† These investigations establish the existence of manifolds which do not generate projections. We show that we need never consider such manifolds if they are avoided at the outset, for the operations we perform do not lead to them.

This paper treats first the subject of projections in spaces of a rather general type. The reflexive property (see definition below)‡ is assumed in order to insure the existence of a limit for monotone sequences of projections. This leads to the establishment of the existence of least upper and greatest lower bounds of sets of permutable projections. Subsequently, a calculus is developed for operators which are defined by means of a resolution of the identity. This calculus possesses properties as extensive as those found in

---

\* Presented to the Society, February 26, 1938; received by the editors December 31, 1937.

† F. J. Murray, *On complementary manifolds and projections in the spaces  $L_p$  and  $l_p$* , these Transactions, vol. 41 (1937), pp. 138-152.

‡ We use the word "reflexive" in preference to "regular" which was introduced by H. Hahn, *Über lineare Gleichungssysteme in linearen Räumen*, Journal für die reine und angewandte Mathematik, vol. 157, pp. 214-229.

Hilbert space.\* The problem of characterizing in simple fashion operators which may be resolved is still open. It may be of interest to note that preliminary attempts in that direction incline one to an optimistic outlook.

## II. THE ALGEBRA OF PROJECTIONS

1. **The space.** We operate in a normed linear vector space  $\mathfrak{B}$  whose elements will be designated by  $f, g, h, \dots$ . Addition of elements and multiplication by complex numbers (denoted by  $\rho, \sigma, \tau, \dots$ ) is permitted subject to the customary restrictions. The norm of  $f$ ,  $\|f\|$ , is a real-valued function which satisfies the conditions  $\|f\| \geq 0$ ,  $\|f\| = 0$  implies  $f = 0$ ,  $\|\rho f\| = |\rho| \|f\|$ , and  $\|f + g\| \leq \|f\| + \|g\|$ . If we write distance  $(f, g) = \|f - g\|$ , the norm metrizes  $\mathfrak{B}$ . If the sequence  $\{f_n\}$  converges to  $f$  in this metric, we often write  $f_n \rightarrow f$  for  $\|f - f_n\| \rightarrow 0$ . The space  $\mathfrak{B}$  is assumed to be complete in this metric.†

An operation  $O$  is a function whose domain is the space  $\mathfrak{B}$  and whose range is a subset of a space of the same type  $\mathfrak{B}_1$ .  $O$  is said to be distributive if  $O(\rho f + \sigma g) = \rho Of + \sigma Og$ . If  $O$  is distributive and continuous at every point of  $\mathfrak{B}$ , then  $O$  is said to be linear. It is known that if  $O$  is linear (and only then if  $O$  is distributive), there exists a constant  $K \geq 0$  independent of  $f \in \mathfrak{B}$ , such that  $\|Of\| \leq K\|f\|$ . The least such constant  $K$  is called the bound of  $O$ . The operation  $O$  is said to be closed if  $f_n \rightarrow f$ ,  $Of_n \rightarrow g$  implies  $Of = g$ . A closed distributive operation is linear.

An operation whose range is contained in the initial space  $\mathfrak{B}$  is called an operator or transformation. Operators are denoted by  $A, B, P, Q, \dots$ . Two special operators and their defining equations are  $0, I$ , with  $0f = 0$ ,  $If = f$ . The bounds of  $A, P, \dots$  are written  $|A|, |P|, \dots$ .

An operation whose range is a set of complex numbers is called a functional. Functionals will be denoted by  $F, G, \dots$ . The totality of linear functionals defined on  $\mathfrak{B}$  is a linear set. If we write  $\|F\|$  for the bound of  $F$ , the set of functionals is a complete space  $(\mathfrak{B})$  of the same type as  $\mathfrak{B}$ . The space  $(\mathfrak{B})$  is said to be the space adjoint to  $\mathfrak{B}$ . The space  $((\mathfrak{B}))$  may now be introduced in evident fashion.

If  $A$  is a linear operator, and  $F$  is a linear functional, both defined on  $\mathfrak{B}$ , then  $F(A)$  is a linear functional. Indeed,  $F(A[f + g]) = F(Af + Ag) = F(Af) + F(Ag)$ . Also

\* For recent developments in the theory of the operational calculus in Hilbert space cf. J. von Neumann, *Über Funktionen von Funktionaloperatoren*, Annals of Mathematics, (2), vol. 32 (1931), pp. 191-226; M. H. Stone, *Linear Transformations in Hilbert Space and their Applications to Analysis*, American Mathematical Society Colloquium Publications, vol. 15, New York, 1932, chap. 6.

† These spaces form the subject of Banach's treatise *Théorie des Opérations Linéaires*. We shall make free and constant use of results there discussed. That these results are for the most part valid in complex spaces has been pointed out recently by various authors.

$$|F(Af)| \leq \|F\| \cdot \|Af\| \leq \|F\| \cdot \|A\| \cdot \|f\|.$$

The correspondence  $F$  to  $F(A)$  defined over  $(\mathfrak{B})$  is linear; for it is distributive and we have just seen that it is bounded. This linear correspondence is written in the operator form,  $G = \bar{A}(F)$ . It is easy to prove that  $|\bar{A}| = |A|$ ,  $(\bar{A} + \bar{B}) = \overline{A+B}$ ,  $(\bar{A} \cdot \bar{B}) = \overline{B \cdot A}$ . The operator  $\bar{A}$  is called the adjoint of  $A$ .

Consider now the spaces  $\mathfrak{B}$ ,  $(\mathfrak{B})$ ,  $((\mathfrak{B}))$ . If  $f \in \mathfrak{B}$  is fixed and  $F \in (\mathfrak{B})$  is variable,  $\Phi(F) = F(f)$  is an element of  $((\mathfrak{B}))$ . For  $\Phi$  is distributive and  $|\Phi(F)| = |F(f)| \leq \|f\| \cdot \|F\|$ ; hence  $\Phi$  is linear. From this it also follows that  $\|\Phi\| \leq \|f\|$ . Since a  $G \in (\mathfrak{B})$  exists such that  $Gf = \|f\|$ ,  $\|G\| = 1$ , we have  $|\Phi(G)| = |G(f)| = \|f\| \leq \|\Phi\| \cdot \|G\| = \|\Phi\|$ ; we must conclude that  $\|\Phi\| = \|f\|$ . The correspondence  $f$  to  $\Phi$  is a linear isometric map of  $\mathfrak{B}$  on a subset of  $((\mathfrak{B}))$ . A space is said to be reflexive if the range of this correspondence is  $((\mathfrak{B}))$  in its entirety. If  $\mathfrak{B}$  is reflexive, we write, for short,  $((\mathfrak{B})) = \mathfrak{B}$ .

*We assume that the space  $\mathfrak{B}$  is reflexive.*

**2. Manifolds.** A set  $M$  of elements is said to be linear if, when  $f, g \in M$ ,  $f + g \in M$ ,  $\rho f \in M$ ,  $\rho$  a complex number. If  $M$  is linear and closed, it is called a closed linear manifold, or for short, a manifold. Manifolds will be designated by the letters  $\mathfrak{M}$ ,  $\mathfrak{N}$ ,  $\dots$ . Let  $\{\mathfrak{M}_\alpha\}$  be any set of closed linear manifolds. Then there exists a smallest manifold  $\mathfrak{M}$  containing each  $\mathfrak{M}_\alpha$ ; we denote this by writing  $\mathfrak{M} = \sum_\alpha^* \mathfrak{M}_\alpha$  (or, in case the index  $\alpha$  ranges over two elements only,  $\mathfrak{M} = \mathfrak{M}_1 + \mathfrak{M}_2$ ). The largest manifold  $\mathfrak{N}$  contained in each  $\mathfrak{M}_\alpha$  is precisely the set intersection of the  $\mathfrak{M}_\alpha$ ,  $\mathfrak{N} = \prod_\alpha \mathfrak{M}_\alpha$  (or, as above,  $\mathfrak{N} = \mathfrak{M}_1 \cdot \mathfrak{M}_2$ ).

The elements  $f \in \mathfrak{B}$ ,  $F \in (\mathfrak{B})$ , are said to be orthogonal to each other if  $F(f) = 0$ . If  $\mathfrak{M} \subset \mathfrak{B}$  is any manifold, the set of all elements  $F \in (\mathfrak{B})$  orthogonal to each element of  $\mathfrak{M}$  is a closed linear manifold  $(\mathfrak{N})$ . Such a manifold  $(\mathfrak{N})$  is called the orthogonal complement of  $\mathfrak{M}$  and is denoted by  $\mathfrak{M}^\perp$ . If each element of  $\mathfrak{M}$  is orthogonal to each element of  $(\mathfrak{N})$ , we write  $\mathfrak{M} \perp (\mathfrak{N})$ . If  $(\mathfrak{N})$  is any manifold in  $(\mathfrak{B})$ , by the orthogonal complement  $\mathfrak{N}$  of  $(\mathfrak{N})$ , we mean the totality of elements in  $\mathfrak{B}$  orthogonal to  $(\mathfrak{N})$ . We note that for  $\mathfrak{N} \subset \mathfrak{B}$ ,  $\mathfrak{N}^{\perp\perp} = \mathfrak{N}$ ; here  $\mathfrak{N}^{\perp\perp}$  means  $\mathfrak{N}^\perp$  where  $\mathfrak{N} = \mathfrak{N}^\perp$ . For clearly  $\mathfrak{N}^{\perp\perp} \supset \mathfrak{N}$ . Let us assume that  $f \in \mathfrak{N}^{\perp\perp}$ ,  $f \notin \mathfrak{N}$ . Then there exists an  $F \in (\mathfrak{B})$  such that  $Ff = 1$ ,  $F \perp \mathfrak{N}$ . Thus  $F \in \mathfrak{N}^\perp$ ; hence  $Ff = 0$ . We conclude that  $\mathfrak{N}^{\perp\perp} = \mathfrak{N}$ . If  $(\mathfrak{N}) \subset (\mathfrak{B})$ , then clearly  $(\mathfrak{N})^{\perp\perp} \supset (\mathfrak{N})$ . To establish the equality of these manifolds, we rely on the reflexive character of the space. Assume  $F \in (\mathfrak{N})^{\perp\perp}$ ,  $F \notin (\mathfrak{N})$ . Then there exists an  $f \in \mathfrak{B}$  such that  $f \perp (\mathfrak{N})$ ,  $F(f) = 1$ ; but this means that  $f \in (\mathfrak{N})^\perp$  and  $Ff = 0 \neq 1$ . Thus  $(\mathfrak{N})^{\perp\perp} = (\mathfrak{N})$ .

We terminate this section with a definition and theorem centering about the operation  $\dot{+}$ . Let  $\mathfrak{M}$  and  $\mathfrak{N}$  be two closed linear manifolds which have the property that there exists a constant  $k > 0$  such that for every  $f \in \mathfrak{M}$  and  $g \in \mathfrak{N}$ ,

$\|f+g\| \geq k\|f\|$ . Then there exists a constant  $k' > 0$  such that  $\|f+g\| \geq k'\|g\|$ . In fact, the inequality  $\|f+g\| \geq \|g\| - \|f\|$  together with the assumed inequality yields  $(1+k)\|f+g\| \geq k\|g\|$  which is what we desire with  $k'$  equal to  $k/(1+k)$ . The symmetry in the roles of  $\mathfrak{M}$  and  $\mathfrak{N}$  allows us to frame the following definition:

**DEFINITION 1.** Two closed linear manifolds  $\mathfrak{M}$  and  $\mathfrak{N}$  will be said to be disjoint if there exists a constant  $k > 0$  such that for every  $f \in \mathfrak{M}$  and for every  $g \in \mathfrak{N}$ ,  $\|f+g\| \geq k\|f\|$ .

**THEOREM 2.1.** Two closed linear manifolds  $\mathfrak{M}$  and  $\mathfrak{N}$  are disjoint if and only if they satisfy the following conditions:

- (1) The manifolds have only the element zero in common.
- (2) The set of all elements of the form  $f+g$ ,  $f \in \mathfrak{M}$ ,  $g \in \mathfrak{N}$ , is a closed linear manifold.

We assume  $\mathfrak{M}$  and  $\mathfrak{N}$  disjoint. Let  $f \in \mathfrak{M} \cap \mathfrak{N}$ . Then  $-f \in \mathfrak{N}$ , and by the foregoing definition  $0 = \|f-f\| \geq k\|f\|$  with  $k > 0$ . Hence  $\|f\| = 0$ ,  $f = 0$ . This establishes (1). In proving (2), we note that the set of elements of the form  $f+g$  is linear. Suppose  $h_n = f_n + g_n$ ,  $(n = 1, 2, \dots)$ , and that  $h_n \rightarrow h$ . We have  $\|h_n - h_m\| \geq k\|f_n - f_m\|$ ; since  $\|h_n - h_m\| \rightarrow 0$ ,  $\|f_n - f_m\| \rightarrow 0$ . The sequence  $\{f_n\}$  is convergent to an element  $f$  which is in  $\mathfrak{M}$  since  $\mathfrak{M}$  is closed. This implies  $g_n \rightarrow g \in \mathfrak{N}$ . Thus  $h = f+g$  and (2) is established.

We now assume conditions (1) and (2). The manifold of all elements of the form  $f+g$  is a complete linear space  $\mathfrak{E}$  of the same type as  $\mathfrak{B}$ . (1) implies directly that all elements in  $\mathfrak{E}$  can be expressed in the form  $f+g$  in only one way. The operator  $A$  which carries  $f+g$  into  $f$ ,  $A(f+g) = f$ , is distributive. Furthermore the conditions  $h_n = f_n + g_n$ ,  $(n = 1, 2, \dots)$ ,  $h_n \rightarrow h$ ,  $f_n \rightarrow f$  in  $\mathfrak{M}$  imply  $g_n \rightarrow g$  in  $\mathfrak{N}$ , where  $h = f+g$ . Thus the operator  $A$  is closed. It is therefore linear. Choose  $k > 0$  so that  $\|Ah\| \leq \|h\|/k$ . Then  $k\|A(f+g)\| = k\|f\| \leq \|f+g\|$ . This terminates the proof.

**3. Projections.** This section is devoted to the development of an elementary theory of projections in  $\mathfrak{B}$ .

**DEFINITION 2.** A linear operator  $P$  is called a projection if  $P^2 = P$ .

**THEOREM 2.2.**<sup>†</sup> Let  $P$  be any projection in  $\mathfrak{B}$ ; let  $\mathfrak{M}$  be the set of elements  $f$  for which  $Pf = f$ ; let  $\mathfrak{N}$  be the set of elements  $g$  for which  $Pg = 0$ . Then  $\mathfrak{M}$  and  $\mathfrak{N}$  are disjoint closed linear manifolds and  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ . Conversely, if  $\mathfrak{M}$  and  $\mathfrak{N}$  are disjoint closed linear manifolds for which  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ , there exists a unique projection  $P$  which satisfies the equations  $Pf = f$ ,  $f \in \mathfrak{M}$ ;  $Pg = 0$ ,  $g \in \mathfrak{N}$ .

<sup>†</sup> The proof of this theorem resembles closely that of Lemma 1.1.1, p. 138, given by Murray loc. cit.

Let  $P$  be a projection. We note that if  $f, g \in \mathfrak{M}$ ,  $pf, f+g \in \mathfrak{M}$ . Furthermore, since  $P$  is continuous,  $\mathfrak{M}$  is closed; thus  $\mathfrak{M}$  is a manifold. Similarly,  $\mathfrak{N}$  is a manifold. If  $f \in \mathfrak{B}$ , then  $f = Pf + (f - Pf)$ . Since  $P^2 = P$ ,  $Pf \in \mathfrak{M}$ ,  $f - Pf \in \mathfrak{N}$ . Thus  $\mathfrak{M} + \mathfrak{N} = \mathfrak{B}$ . Since  $\|Pf\| \leq |P| \cdot \|f\| = |P| \cdot \|Pf + (f - Pf)\|$ ,  $\mathfrak{M}$  and  $\mathfrak{N}$  are disjoint.

We turn to the converse. Let  $\mathfrak{M}$  and  $\mathfrak{N}$  be disjoint,  $\mathfrak{M} + \mathfrak{N} = \mathfrak{B}$ . Let  $h \in \mathfrak{B}$ ,  $h = f + g$ ,  $f \in \mathfrak{M}$ ,  $g \in \mathfrak{N}$ . Then the operator  $P$  for which  $Ph = f$  is distributive. As in Theorem 2.1,  $P$  is closed, hence linear. Furthermore,  $P^2h = P(Ph) = Pf = f = Ph$ . Thus  $P$  has the properties required by the theorem. Any linear operator which is identical with  $P$  on  $\mathfrak{M}$  and on  $\mathfrak{N}$  is identical with  $P$  in  $\mathfrak{B}$  since  $\mathfrak{M} + \mathfrak{N} = \mathfrak{B}$ .

The manifolds  $\mathfrak{M}$  and  $\mathfrak{N}$  described above are said to be associated to the projection  $P$ . We shall sometimes denote the manifolds associated to  $P_1$  and  $P_2, \dots$  by  $\mathfrak{M}_{P_1}, \mathfrak{N}_{P_1}, \dots$ , also by  $\mathfrak{M}_1, \mathfrak{N}_1, \dots$ .

**THEOREM 2.3.** *If  $P$  is a projection in  $\mathfrak{B}$ , then  $\bar{P}$  is a projection in  $(\mathfrak{B})$ . If  $\mathfrak{M}, \mathfrak{N}$  and  $(\mathfrak{M}), (\mathfrak{N})$  are the manifolds associated to  $P$  and  $\bar{P}$ , respectively, then  $(\mathfrak{M}) = \mathfrak{N}^\perp$  and  $(\mathfrak{N}) = \mathfrak{M}^\perp$ .*

The relation  $P^2 = P$  implies  $(\bar{P} \cdot \bar{P}) = \bar{P}^2 = \bar{P}$ ; thus  $\bar{P}$  is a projection.  $\bar{P}$  is defined by the equations  $\bar{P}(F) = G$  where  $Gf = F(Pf)$ . Let  $G \in (\mathfrak{M})$ . Then  $G = \bar{P}(G)$  and  $Gf = G(Pf) = 0$  if  $f \in \mathfrak{N}$ . Thus  $(\mathfrak{M}) \perp \mathfrak{N}$  or  $(\mathfrak{M}) \subset \mathfrak{N}^\perp$ . Now let  $G$  be any element orthogonal to  $\mathfrak{N}$ . We shall show that  $G = \bar{P}(G)$ . Let  $f$  be arbitrary in  $\mathfrak{B}$ . Then  $Gf = G(Pf + (f - Pf)) = GPf + G(f - Pf) = G(Pf)$  since  $f - Pf \in \mathfrak{N}$ . Thus  $(\mathfrak{M}) = \mathfrak{N}^\perp$ . The proof that  $(\mathfrak{N}) = \mathfrak{M}^\perp$  is similar.

**THEOREM 2.4.** *If  $P_1$  and  $P_2$  are projections and  $\mathfrak{M}_1, \mathfrak{N}_1, \mathfrak{M}_2, \mathfrak{N}_2$  are their associated manifolds, then*

- (1)  $P_1 \cdot P_2$  is a projection if and only if  $P_1(\mathfrak{M}_2) \subset \mathfrak{M}_2 + \mathfrak{N}_1\mathfrak{M}_2$ ;
- (2)  $P_1 + P_2$  is a projection if and only if  $P_1P_2 = P_2P_1 = 0$ .

(1) Assume  $P_1P_2$  a projection. Let  $f \in \mathfrak{M}_2$ ,  $P_1f = g + h$ ,  $g \in \mathfrak{M}_2$ ,  $h \in \mathfrak{N}_2$ . Then  $P_1f = P_1P_2f = P_1P_2P_1P_2f = P_1P_2P_1f = P_1g$ . Thus  $P_1(f - g) = 0$ ,  $f - g = f - P_1f + h \in \mathfrak{N}_1$ , or  $h \in \mathfrak{N}_1$ . This establishes that  $P_1(\mathfrak{M}_2) \subset \mathfrak{M}_2 + \mathfrak{N}_1\mathfrak{M}_2$ .

Now assume that  $P_1(\mathfrak{M}_2) \subset \mathfrak{M}_2 + \mathfrak{N}_1\mathfrak{M}_2$ . Note that  $P_1P_2$  is distributive and bounded, hence linear. We show that  $(P_1P_2)^2 = P_1P_2$ . The equation holds on  $\mathfrak{N}_2$ . It will suffice to establish it on  $\mathfrak{M}_2$ . For  $f \in \mathfrak{M}_2$ ,  $P_1f = g + h$ ,  $g \in \mathfrak{M}_2$ ,  $h \in \mathfrak{N}_1\mathfrak{M}_2$ ,<sup>†</sup>  $P_1P_2P_1P_2f = P_1P_2P_1f = P_1g = P_1(g + h) = P_1f = P_1P_2f$ . This completes the proof of part (1).

<sup>†</sup> The manifolds  $\mathfrak{M}_2$  and  $\mathfrak{N}_2$  are disjoint. Thus  $\mathfrak{M}_2$  and  $\mathfrak{N}_1\mathfrak{M}_2$  are disjoint. By Theorem 2.1, any element in  $\mathfrak{M}_2 + \mathfrak{N}_1\mathfrak{M}_2$  has the form  $g + h$  which we describe above.

(2) We note that  $(P_1 + P_2)^2 = P_1 + P_2 + P_1P_2 + P_2P_1$ . If  $P_1P_2 = P_2P_1 = 0$ ,  $P_1 + P_2$  is a projection. Assume now that  $P_1 + P_2$  is a projection. Then  $P_1P_2 + P_2P_1 = 0$ . We shall see that  $P_2P_1 = 0$ . Note that  $P_2P_1 = 0$  if  $f \in \mathfrak{M}_1$ . If  $f \in \mathfrak{M}_1$ ,  $P_1P_2f + P_2P_1f = P_1P_2f + P_2f = 0$ . Let  $P_2f = g + h$ ,  $g \in \mathfrak{M}_1$ ,  $h \in \mathfrak{M}_1$ , then  $P_1P_2f + P_2f = g + g + h = 0$ ,  $2g = -h$ ,  $g = h = 0$ ,  $P_2f = P_2P_1f = 0$ .

If for the projections  $P_1$  and  $P_2$ ,  $\mathfrak{M}_1 \supset \mathfrak{M}_2$ ,  $\mathfrak{M}_1 \subset \mathfrak{M}_2$ , we write  $P_1 > P_2$ . If  $P_1 > P_2$  and  $P_2 > P_3$ , then  $P_1 > P_3$ . If  $P_1 > P_2$ , then  $\overline{P_1} > \overline{P_2}$ . If  $P_1 > P_2$ , then  $P_1P_2 = P_2P_1 = P_2$ . For any projection  $P$ ,  $I > P > 0$ .

**THEOREM 2.5.** *If  $P_1$  and  $P_2$  are permutable projections, then  $P_1P_2$  and  $Q = P_1 + P_2 - P_1P_2$  are projections. The associated manifolds of these projections are related by the equations  $\mathfrak{M}_{P_1P_2} = \mathfrak{M}_{P_1} \cdot \mathfrak{M}_{P_2}$ ,  $\mathfrak{N}_{P_1P_2} = \mathfrak{N}_{P_1} \dot{+} \mathfrak{N}_{P_2}$ ;  $\mathfrak{M}_Q = \mathfrak{M}_{P_1} \dot{+} \mathfrak{M}_{P_2}$ ,  $\mathfrak{N}_Q = \mathfrak{N}_{P_1} \cdot \mathfrak{N}_{P_2}$ . If  $P_1 > P_2$ , then  $R = P_1 - P_2$  is a projection and  $\mathfrak{M}_R = \mathfrak{M}_{P_1} \cdot \mathfrak{N}_{P_2}$ ,  $\mathfrak{N}_R = \mathfrak{M}_{P_2} \dot{+} \mathfrak{N}_{P_1}$ .*

Direct computation yields  $(P_1P_2)^2 = P_1P_2$ ,  $Q^2 = Q$ , and if  $P_1 > P_2$ ,  $(P_1 - P_2)^2 = P_1 - P_2$ . Clearly,  $\mathfrak{M}_{P_1P_2} \supset \mathfrak{M}_{P_1} \cdot \mathfrak{M}_{P_2}$ . Now let  $P_1P_2f = f$ . Then  $f = P_1P_2f = P_1P_2^2f = P_2P_1P_2f = P_2f$ ; thus  $f \in \mathfrak{M}_2$ . Similarly  $f \in \mathfrak{M}_1$ . Therefore  $\mathfrak{M}_{P_1P_2} = \mathfrak{M}_{P_1} \cdot \mathfrak{M}_{P_2}$ .

Next,  $\mathfrak{M}_{P_1P_2} \supset \mathfrak{N}_{P_1}$ ,  $\mathfrak{M}_{P_1P_2} \supset \mathfrak{N}_{P_2}$ ; hence  $\mathfrak{M}_{P_1P_2} \supset \mathfrak{N}_{P_1} \dot{+} \mathfrak{N}_{P_2}$ . Let  $P_1P_2f = 0$ ,  $f = g + h$ ,  $g \in \mathfrak{M}_2$ ,  $h \in \mathfrak{N}_2$ . Then  $P_1P_2f = P_1g = 0$ ,  $g \in \mathfrak{M}_1$ , and  $f = g + h \in \mathfrak{M}_1 \dot{+} \mathfrak{N}_2$ . Hence  $\mathfrak{N}_{P_1P_2} = \mathfrak{N}_{P_1} \dot{+} \mathfrak{N}_{P_2}$ .

As for  $Q$ , let  $f \in \mathfrak{M}_{P_1}$ ,  $g \in \mathfrak{M}_{P_2}$ , then  $Q(f + g) = f + P_1g + P_2f + g - P_1P_2f - P_1P_2g = f + g$  since  $P_1g + P_2f = P_1P_2g + P_1P_2f$ . Hence  $\mathfrak{M}_Q \supset \mathfrak{M}_{P_1} \dot{+} \mathfrak{M}_{P_2}$ . If  $Qf = f$ , write  $f = g + h$ ,  $g \in \mathfrak{M}_1$ ,  $h \in \mathfrak{N}_1$ ,  $Qf = g + P_2g + P_2h - P_2P_1(g + h) = g + P_2g + P_2h - P_2g = g + h$ ;  $P_2h = h$ ,  $h \in \mathfrak{M}_2$ , and  $f \in \mathfrak{M}_1 \dot{+} \mathfrak{M}_2$ .

Next, note that  $\mathfrak{N}_Q \supset \mathfrak{N}_{P_1} \cdot \mathfrak{N}_{P_2}$ . If  $f \in \mathfrak{N}_Q$ , then  $Qf = (P_1 + P_2 - P_1P_2)f = 0$ ,  $P_1f = (P_1 - I)P_2f = P_2^2f = (P_1 - I)P_1P_2f = 0$ , and  $f \in \mathfrak{N}_1$ ; likewise  $f \in \mathfrak{N}_2$ . Hence  $\mathfrak{N}_Q = \mathfrak{N}_{P_1} \cdot \mathfrak{N}_{P_2}$ .

We consider  $R$ . Clearly,  $\mathfrak{M}_R \supset \mathfrak{M}_{P_1} \cdot \mathfrak{N}_{P_2}$ . If  $(P_1 - P_2)f = f$ ,  $P_2f = P_2(P_1 - P_2)f = 0$ ,  $f \in \mathfrak{N}_{P_2}$ , and hence  $f \in \mathfrak{M}_{P_1}$ . Note that  $R = 0$  on  $\mathfrak{M}_{P_2}$  and on  $\mathfrak{N}_{P_1}$ . And if  $Rf = 0$ , then  $P_1f = P_2f$ ,  $f = P_2f + (f - P_2f) = P_2f + (f - P_1f)$ ; and since  $P_2f \in \mathfrak{M}_{P_2}$ ,  $f - P_1f \in \mathfrak{N}_{P_1}$ , we must have  $\mathfrak{N}_R = \mathfrak{M}_{P_2} \dot{+} \mathfrak{N}_{P_1}$ .

### III. INFINITE SYSTEMS OF PROJECTIONS

In the beginning of this chapter, monotone sequences of projections are treated. We determine conditions under which a limit operator exists. Subsequently, the notion of least upper bound and greatest lower bound of sets of permutable projections is examined.

We remark first that if  $\mathfrak{M}$  and  $\mathfrak{N}$  are disjoint, and if there exists an  $f (\neq 0) \in \mathfrak{B}$  not in  $\mathfrak{M} \dot{+} \mathfrak{N}$ , then there exists an  $F (\neq 0) \in \mathfrak{M} \cdot \mathfrak{N}^\perp$ . Thus if  $\mathfrak{M}^\perp$

and  $\mathfrak{M}^\perp$  are disjoint,  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ . In proof, let  $f \in \mathfrak{M} \dot{+} \mathfrak{N}$ , ( $f \neq 0$ ); then there exists an  $F \in (\mathfrak{B})$  such that  $Ff = 1$ ,  $F \perp \mathfrak{M} \dot{+} \mathfrak{N}$ . Hence we have  $F (\neq 0) \in \mathfrak{M}^\perp$ , and  $F \in \mathfrak{N}^\perp$ .

**THEOREM 3.1.** *Let  $\{P_n\}$  be a sequence of projections for which  $P_n < P_{n+1}$ ,  $|P_n| \leq K$ , ( $n = 1, 2, \dots$ ). Let the adjoint of  $P_n$  be  $\bar{P}_n$ , and let  $\mathfrak{M}_n, \mathfrak{N}_n, (\mathfrak{M})_n, (\mathfrak{N})_n$  denote the manifolds associated to  $P_n$  and  $\bar{P}_n$ . Let  $\mathfrak{M} = \sum_1^\infty \mathfrak{M}_\alpha$ ,  $\mathfrak{N} = \prod_1^\infty \mathfrak{N}_\alpha$ ,  $(\mathfrak{M}) = \sum_1^\infty (\mathfrak{M})_\alpha$ ,  $(\mathfrak{N}) = \prod_1^\infty (\mathfrak{N})_\alpha$ . Then*

- (1)  $\mathfrak{M}$  and  $\mathfrak{N}$  are disjoint,  $(\mathfrak{M})$  and  $(\mathfrak{N})$  are disjoint;
- (2)  $(\mathfrak{M}) = \mathfrak{M}^\perp$ ,  $(\mathfrak{N}) = \mathfrak{M}^\perp$ ;  $\mathfrak{M} = (\mathfrak{N})^\perp$ ,  $\mathfrak{N} = (\mathfrak{M})^\perp$ ;
- (3)  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ ;  $(\mathfrak{M}) \dot{+} (\mathfrak{N}) = (\mathfrak{B})$ .

(1) Let  $f \in \mathfrak{M}$ ,  $g \in \mathfrak{N}$ . Then there exist elements  $f_n \in \mathfrak{M}_n$  such that  $f_n \rightarrow f$ . Thus  $\|P_n(f_n + g)\| = \|f_n\| \leq K\|f_n + g\|$ . As the case  $|P_n| = 0$ , ( $n = 1, 2, \dots$ ), is trivial, we assume  $K > 0$ . Thus  $\|f_n + g\| \geq \|f_n\|/K$ ; therefore  $\|f + g\| \geq \|f\|/K$  which implies the disjointness of  $\mathfrak{M}$  and  $\mathfrak{N}$ . To show the same for  $(\mathfrak{M})$  and  $(\mathfrak{N})$ , we observe that  $|\bar{P}_n| = |P_n| \leq K$ ,  $\bar{P}_n < P_{n+1}$ , ( $n = 1, 2, \dots$ ), and apply the result just obtained.

(2) Since  $(\mathfrak{N})_n = \mathfrak{M}_n^\perp$  by Theorem 2.3,  $\mathfrak{M}_n \perp (\mathfrak{N})_n$ , ( $n = 1, 2, \dots$ ). Hence,  $(\mathfrak{N}) \perp \sum_1^\infty \mathfrak{M}_\alpha = \mathfrak{M}$ , or  $(\mathfrak{N}) \subset \mathfrak{M}^\perp$ . Now let  $F \in \mathfrak{M}^\perp$ . Then  $F \perp \mathfrak{M}_n$ ; hence  $F \in (\mathfrak{M})_n$ , ( $n = 1, 2, \dots$ ),  $F \in \prod_1^\infty (\mathfrak{M})_\alpha = (\mathfrak{M})$ . Thus  $(\mathfrak{N}) = \mathfrak{M}^\perp$ .

Since  $(\mathfrak{M})_n \perp \mathfrak{N}_n$ ,  $(\mathfrak{M})_n \perp \mathfrak{N}$ , we have  $(\mathfrak{M}) = \sum_1^\infty (\mathfrak{M})_\alpha \perp \mathfrak{N}$  and  $(\mathfrak{M})^\perp \supset \mathfrak{N}$ . Let  $f \in (\mathfrak{M})^\perp$ , that is, let  $f \in \mathfrak{B}$ ,  $f \perp (\mathfrak{M})$ . Then  $f \perp (\mathfrak{M})_n$ ,  $f \in \mathfrak{N}_n$ ,  $f \in \mathfrak{N}$  (see the discussion under II, 2). Hence  $(\mathfrak{M})^\perp = \mathfrak{N}$ .

Since  $(\mathfrak{N}) = \mathfrak{M}^\perp$ ,  $(\mathfrak{N})^\perp = \mathfrak{M}^{\perp\perp} = \mathfrak{M}$ . It remains to show that  $(\mathfrak{M}) = \mathfrak{N}^\perp$ . Starting with  $(\mathfrak{M})^\perp = \mathfrak{N}$ , we obtain  $(\mathfrak{M})^{\perp\perp} = \mathfrak{N}^\perp$ . Since  $\mathfrak{B}$  is reflexive,  $(\mathfrak{M})^{\perp\perp} = (\mathfrak{M})$  and  $(\mathfrak{M}) = \mathfrak{N}^\perp$ .

(3) By the remark preceding this theorem and by (1) and (2),  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ . Similarly,  $(\mathfrak{M}) \dot{+} (\mathfrak{N}) = (\mathfrak{B})$ .

**THEOREM 3.2.** *Let  $\{P_n\}$  be a sequence of projections for which  $|P_n| \leq K$ ,  $P_n < P_{n+1}$ , ( $n = 1, 2, \dots$ ). Then there exists a projection  $P$  having the following properties:*

- (1)  $\mathfrak{M}_P = \sum_1^\infty \mathfrak{M}_\alpha$ ,  $\mathfrak{N}_P = \prod_1^\infty \mathfrak{N}_\alpha$ .
- (2)  $|P| \leq K$ .
- (3) For any  $f \in \mathfrak{B}$ ,  $\|(P - P_n)f\| \rightarrow 0$ .
- (4)  $P > P_n$ , ( $n = 1, 2, \dots$ ). If  $Q$  is a projection such that  $Q > P_n$ , ( $n = 1, 2, \dots$ ), then  $Q > P$ .
- (5) If  $P_n$  is permutable with a linear operator  $A$ , ( $n = 1, 2, \dots$ ), then  $P$  is permutable with  $A$ .

Similar conclusions may be drawn for sequences  $\{P_n\}$  in which the hypothesis  $P_n > P_{n+1}$  replaces  $P_n < P_{n+1}$ , ( $n = 1, 2, \dots$ ).

(1) We define  $P$  to be the projection whose associated manifolds are  $\mathfrak{M}_P = \sum_1^* \mathfrak{M}_a$ ,  $\mathfrak{N}_P = \prod_1 \mathfrak{N}_a$ . Since these manifolds are disjoint and since  $\mathfrak{M}_P \dot{+} \mathfrak{N}_P = \mathfrak{B}$  (by the previous theorem),  $P$  is uniquely defined (Theorem 2.3).

(2) That  $|P| \leq K$  is apparent from the proof of the first statement in Theorem 3.1.

(3) Assume that  $f \in \mathfrak{B}$ . Then  $Pf \in \mathfrak{M}_P$ , and there exist elements  $g_n \in \mathfrak{M}_n$ , ( $n=1, 2, \dots$ ), such that  $g_n \rightarrow Pf$ . Now  $Pf - P_n f = Pf - g_n + P_n(g_n + f - Pf) - P_n f$  and  $\|Pf - P_n f\| \leq \|Pf - g_n\| + \|P_n(g_n + f - Pf)\| \rightarrow 0$  since  $|P_n| \leq K$ .

(4) That  $P > P_n$  is clear from (1). If  $Q > P_n$ , then  $\mathfrak{M}_Q \supset \mathfrak{M}_n$ , ( $n=1, 2, \dots$ ), hence  $\mathfrak{M}_Q \supset \mathfrak{M}_P$ . Similarly,  $\mathfrak{N}_Q \subset \mathfrak{N}_P$ . Thus  $P < Q$ .

(5) Since  $P_n A f = A P_n f$ ,  $P_n A f \rightarrow P A f$ ,  $A P_n f \rightarrow A P f$ , then  $P A f = A P f$ .

The proof of the last statement in the theorem presents no difficulties.

Let  $P_1, P_2, \dots, P_n$  be mutually permutable projections. Then  $Q = \prod_1^n P_n$  is a projection. It is readily seen that  $\mathfrak{M}_Q = \prod_1^n \mathfrak{M}_a$ ,  $\mathfrak{N}_Q = \sum_1^n \mathfrak{N}_a$ . The manifolds associated with the projection  $I - P_1$  are  $\mathfrak{M}_{I-P_1} = \mathfrak{N}_1$ ,  $\mathfrak{N}_{I-P_1} = \mathfrak{M}_1$ . Thus the manifolds associated with the projection  $R = I - \prod_1^n (I - P_n)$  are  $\mathfrak{M}_R = \sum_1^n \mathfrak{M}_a$ ,  $\mathfrak{N}_R = \prod_1^n \mathfrak{N}_a$ . The projection  $R$  formed in this way is denoted by the symbol  $R = \sum_1^n P_n$  (or  $R = P_1 \dot{+} \dots \dot{+} P_n$ ).

**DEFINITION 3.** A set  $\Omega$  of permutable projections is called a lattice<sup>†</sup> of projections if, when  $P, P_1, P_2 \in \Omega$ , then  $I - P, P_1 P_2 \in \Omega$ . The lattice is said to be  $K$ -bounded if  $|P| \leq K$  for every  $P \in \Omega$ .

If  $P_1, P_2 \in \Omega$ , then  $P_1 \dot{+} P_2 = I - (I - P_1)(I - P_2) \in \Omega$ . Any set  $\mathfrak{M}$  of permutable projections may be embedded in a lattice of projections. Indeed, let  $\mathfrak{P}(\xi_1, \xi_2, \dots, \xi_s)$  represent any polynomial with integral coefficients such that  $\mathfrak{P}(P_1, P_2, \dots, P_s)$  is a projection, for any set of  $s$  mutually permutable projections  $P_1, P_2, \dots, P_s$ .<sup>‡</sup> The set of all projections thus obtained contains  $\mathfrak{M}$  and is a lattice of projections.

**THEOREM 3.3.** Let  $\Omega$  be any  $K$ -bounded lattice of projections, and let  $\{P_a\}$  be any subset of  $\Omega$ . Then the manifolds  $\mathfrak{M} = \sum_a^* \mathfrak{M}_a$  and  $\mathfrak{N} = \prod_a \mathfrak{N}_a$  are disjoint and  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ . Similarly, the manifolds  $\mathfrak{M}' = \prod_a \mathfrak{M}_a$  and  $\mathfrak{N}' = \sum_a^* \mathfrak{N}_a$  are disjoint and  $\mathfrak{M}' \dot{+} \mathfrak{N}' = \mathfrak{B}$ . Let  $P$  be the projection associated to  $\mathfrak{M}$  and  $\mathfrak{N}$ ; let  $P'$  be the projection associated to  $\mathfrak{M}'$  and  $\mathfrak{N}'$ . Then  $|P| \leq K$ ,  $|P'| \leq K$ ,  $P > P_a$ , and  $Q > P_a$  implies  $Q > P$ . Also,  $P_a > P'$ , and  $P_a > Q$  implies  $P' > Q$ . The projections  $P$  and  $P'$  are permutable with any linear operator permutable with  $P_a$ .

Let  $f \in \sum_a^* \mathfrak{M}_a$ ,  $g \in \prod_a \mathfrak{N}_a$ . Then there exist elements  $f_n$  and manifolds  $\mathfrak{M}_r$ , associated to the projections  $P_r \in \Omega$ , ( $n, r=1, 2, \dots$ ), such that  $f_n \in \sum_{r=1}^n \mathfrak{M}_r$ ,

<sup>†</sup> In fact, a lattice of projections constitutes a Boolean algebra.

<sup>‡</sup> In making the substitutions, write  $I$  for  $P^0$ .

$f_n \rightarrow f$ . Let  $Q$  be the projection which is the limit of the monotone sequence  $\{Q_n = \sum_{r=1}^n P_r\}$  (Theorem 3.2). Then  $\mathfrak{M}_Q = \sum_{r=1}^\infty \mathfrak{M}_r \subset \mathfrak{M}$  and  $\mathfrak{N}_Q \supset \mathfrak{N}$ . Since  $|Q| \leq K$  (Theorem 3.2),  $\|Q(f+g)\| = \|f\| \leq K\|f+g\|$ . This proves that  $\mathfrak{M}$  and  $\mathfrak{N}$  are disjoint. Similarly,  $\mathfrak{M}'$  and  $\mathfrak{N}'$  are disjoint.

We prove that  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ . Let  $(\mathfrak{M})_\alpha, (\mathfrak{N})_\alpha$  be the manifolds associated to  $P_\alpha$ . Let  $(\mathfrak{M}) = \sum_\alpha^* (\mathfrak{M})_\alpha, (\mathfrak{N}) = \prod_\alpha (\mathfrak{N})_\alpha$ . Then precisely as in the proof of Theorem 3.1,  $(\mathfrak{M}) = \mathfrak{N}^\perp, (\mathfrak{N}) = \mathfrak{M}^\perp$ ; and since  $(\mathfrak{M})$  and  $(\mathfrak{N})$  are disjoint (by the argument given above)  $\mathfrak{M} \dot{+} \mathfrak{N} = \mathfrak{B}$ . Similarly  $\mathfrak{M}' \dot{+} \mathfrak{N}' = \mathfrak{B}$ .

That  $|P| \leq K, |P'| \leq K$  follows from the inequality  $\|f\| \leq K\|f+g\|$  derived in the first paragraph of this proof. That  $P > P_\alpha, P' < P_\alpha$  is clear. If  $Q > P_\alpha$ , then  $\mathfrak{M}_Q \supset \sum_\alpha^* \mathfrak{M}_\alpha = \mathfrak{M}_P$  and  $\mathfrak{N}_Q \subset \prod_\alpha \mathfrak{N}_\alpha = \mathfrak{N}_P$ ; hence  $Q > P$ . Similarly if  $Q < P_\alpha, Q < P'$ .

We examine the statement on permutability. Let  $AP_\alpha = P_\alpha A, A$  linear, and let  $f \in \mathfrak{B}$ . As in the first paragraph of this proof, we obtain a projection  $Q_1 < P$  such that  $Pf \in \mathfrak{M}_{Q_1}$  or  $Pf = Q_1 f$ . Similarly, we obtain a  $Q_2 < P$  such that  $PAf \in \mathfrak{M}_{Q_2}$  or  $PAf = Q_2 Af$ . By Theorem 3.2,  $Q_1, Q_2$ , and also  $Q = Q_1 \dot{+} Q_2$ , are permutable with  $A$ . We have  $Q_1 f = Q Q_1 f = Q P f = Q f$  and  $Q A f = Q_2 A f$ , and since  $Q A = A Q, A P f = A Q f = A Q f = Q A f = Q_2 A f = P A f$ .

The projections  $P$  and  $P'$  of this theorem will be denoted by the symbols  $\sum_\alpha^* P_\alpha$  and  $\prod_\alpha P_\alpha$ , respectively.

**THEOREM 3.4.** *Let  $\Omega$  be a  $K$ -bounded lattice of projections. Let  $\Omega'$  be the set of all projections of the type  $P = \sum_\alpha^* P_\alpha, P' = \prod_\alpha P_\alpha$  where the sums and products are formed over the subsets of  $\Omega$ . Then  $\Omega'$  may be embedded in a  $K$ -bounded lattice of projections.*

Letting  $P_i$  represent an arbitrary projection in  $\Omega'$ , we create independent variables  $x_i$  in 1-1 correspondence to the  $P_i$ . We define a sequence  $\{M_n\}$  of classes of polynomials in the variables  $x_i$ ;  $M_0$  is the set of all  $x_i$ ; assuming that  $M_n$  is known, we define  $M_{n+1}$ . If  $n$  is even,  $M_{n+1}$  is the set of all polynomials of the form  $y_1, y_2, \dots, y_n$  where  $y_i \in M_n$ . If  $n$  is odd,  $M_{n+1}$  is the set of all polynomials in  $M_n$  to which have been added all polynomials of the form  $1-y$ , where  $y \in M_n$ . This process defines  $M_n$  for  $n=1, 2, \dots$ . Let  $M = \sum_{n=1}^\infty M_n$ . Then  $M$  has the following properties:  $x_i \in M$ ; if  $y, y_1, y_2 \in M$ , then  $1-y, y_1 y_2 \in M$ . In fact, as one may easily see,  $M$  is the smallest set of polynomials possessing these properties. If  $y = y(x_{i_1}, \dots, x_{i_m}) \in M$ , then  $y(P_{i_1}, \dots, P_{i_m})$  is a projection. For there exists a value of  $n$  such that  $y \in M_n$ . If  $n=0$ , our assertion is obvious. A clearly indicated induction settles the case  $n \geq 1$ . Let  $\Gamma$  be the set of all projections  $y(P_{i_1}, \dots, P_{i_m})$ . We shall show that  $\Gamma$  is a  $K$ -bounded lattice of projections.

Since the members of  $\Omega'$  may be permuted among themselves, the same is true of the members of  $\Gamma$ . From the definition of  $M$ , it is clear that  $\Gamma$  is a

lattice of projections; indeed, it is the smallest lattice which includes  $\Omega'$ . We shall see that for arbitrary  $R \in \Gamma$ ,  $f \in \mathfrak{B}$ , there exists a  $Q \in \Omega$  such that  $\|(R-Q)f\|$  is small at will. Then since  $\|Rf\| \leq \|(R-Q)f\| + \|Qf\|$ ,  $\|Rf\| \leq K\|f\|$ .

Let  $P = \sum_{\alpha}^* P_{\alpha} \in \Omega'$ ,  $P_{\alpha} \in \Omega$ ; let  $f \in \mathfrak{B}$ . Then as in Theorem 3.3 there exists a monotone increasing sequence  $\{P_n\}$  such that  $P_n \in \Omega$ ,  $P_n f \rightarrow P f$ . If  $P' = \prod_{\alpha} P_{\alpha} \in \Omega'$ , we may find a decreasing sequence  $\{P'_n\}$ ,  $P'_n \in \Omega$ ,  $P'_n f \rightarrow P' f$ . Now let  $R_1, \dots, R_s \in \Omega'$ ,  $R = \gamma(R_1, \dots, R_s) \in \Gamma$ . Let  $R_{in} f \rightarrow R_i f$ ,  $R_{in} \in \Omega$ , ( $i=1, 2, \dots, s$ ). It is clear from the construction of  $\Gamma$  that  $\gamma(R_{1n}, \dots, R_{sn}) \in \Omega$ . It may also be seen that  $\gamma(R_{1n}, \dots, R_{sn}) f \rightarrow \gamma(R_1, \dots, R_s) f$ .

#### IV. THE THEORY OF PROJECTION MEASURE

In this chapter, we define the notion of a resolution of the identity in  $\mathfrak{B}$ . A theory of projection measure generated by this resolution of the identity is developed. With certain sets of real numbers we associate projections. Products and sums of sets correspond to products and  $*$ -sums of the associated projections.<sup>†</sup>

A set of real numbers  $a < \lambda \leq b$  will be designated by  $\delta$ . Let  $\{\delta_n\}$  be a sequence of such sets; we designate  $\sum_1^{\infty} \delta_n$  by  $\Delta$ . The set  $\Delta$  is said to be a covering of a set  $M$  if  $\Delta \supset M$ .

DEFINITION 4. A set of projections  $E(\lambda)$ ,  $(-\infty < \lambda < \infty)$ , is called a resolution of the identity if

- (1) The projections 0 and  $I$  are in the set;
- (2)  $E(\lambda) > E(\mu)$  for  $\lambda > \mu$ ;
- (3) There exists a constant  $K$  such that for any given real numbers  $a_i, b_i$ , ( $i=1, 2, \dots, n$ ), with  $a_1 \leq b_1 \leq \dots \leq a_i \leq b_i \leq \dots \leq a_n \leq b_n$ , and for any given complex numbers  $\mu_i$ , ( $i=1, 2, \dots, n$ ), with  $|\mu_i| \leq 1$  the bound of the operator

$$\sum_1^n \mu_i [E(b_i) - E(a_i)]$$

does not exceed  $K$ .

(1) and (2) imply the existence of two real numbers  $r, R$  such that  $E(\lambda) = 0$  for  $\lambda \leq r$ ,  $E(\lambda) = I$  for  $\lambda \geq R$ .

We shall prove that if  $\Delta = \sum_1^{\infty} \delta_n$ ,  $\delta_i = \{a_i < \lambda \leq b_i\}$  is any covering of the set of all real numbers  $\lambda$ , and if  $E(\delta_i)$  means  $E(b_i) - E(a_i)$ , then  $\sum_1^{\infty} E(\delta_n) = I$ . Let us write here, as often later,  $E(\Delta)$  for  $\sum_1^{\infty} E(\delta_n)$ . Then  $E(\Delta_n) = \sum_1^n E(\delta_n) \rightarrow E(\Delta)$ . In the first place, for fixed  $f \in \mathfrak{B}$ ,  $F \in \mathfrak{B}$ , the function  $F(E(\lambda)f)$  is of bounded variation and indeed  $\text{var } F(E(\lambda)f) \leq K \cdot |F| \cdot \|f\|$ . For consider a sub-

<sup>†</sup> Such a theory was developed by the author for the case in which  $\mathfrak{B}$  is a (separable) Hilbert space in *Acta Litterarum ac Scientiarum*, vol. 7 (1935), pp. 136-146.

division  $r = a_0 \leq a_1 \leq \dots \leq a_n = R$  of the interval  $\beta = \{r \leq \lambda \leq R\}$  with  $r, R$  as above. Assume

$$F(E(a_j)f) - F(E(a_{j-1})f) = e^{i\theta_j} |F(E(a_j) - E(a_{j-1}))f|, \theta_j \text{ real}, j = 1, 2, \dots, n.$$

Then

$$\begin{aligned} \sum_{j=1}^n |F(E(a_j) - E(a_{j-1}))f| &= \left| \sum_{j=1}^n F(e^{-i\theta_j} [E(a_j) - E(a_{j-1}))f] \right| \\ &= |F(T \cdot f)| \leq |F| \cdot |T| \cdot \|f\|, \end{aligned}$$

$T$  being a certain linear transformation defined by the equation. But  $|T| \leq K$  by Definition 4, (3). Now writing  $\Gamma_n = \beta - \sum_{j=1}^n \delta_{a_j}$ , ( $n = 1, 2, \dots$ ), we see that  $\Gamma_1 \supset \Gamma_2 \supset \dots \supset \Gamma_n \supset \dots$ , and that  $\prod_{n=1}^{\infty} \Gamma_n$  is the null set. Hence

$$|F(f) - FE(\Delta_n)f| = |F(I - E(\Delta_n))f| = |F(E(\Gamma_n)f)| \leq \varlimsup_{\Gamma_n} F(E(\lambda)f).$$

Thus  $F(f) - FE(\Delta_n)f \rightarrow 0$  with  $1/n$ . Since  $E(\Delta_n)f$  converges weakly to  $f$  and strongly to  $E(\Delta)f$ , we have  $E(\Delta)f = f$ , and our proof is complete.

We prove the following theorem:

**THEOREM 4.1.** *Any resolution of the identity may be embedded in a  $K$ -bounded lattice of projections.*

Let  $\delta_1, \dots, \delta_n$  be nonoverlapping intervals. Then  $\sum_{\alpha=1}^n E(\delta_{\alpha})$  is a projection, and its bound does not exceed  $K$  (Definition 4, (3)). The totality  $\Omega$  of projections formed in this manner is a  $K$ -bounded lattice. The members of  $\Omega$  are permutable. If  $P, P_1, P_2 \in \Omega$ , then  $I - P, P_1 P_2 \in \Omega$ , since the complement in the set  $r < \lambda \leq R$  of  $\sum_{\alpha=1}^n \delta_{\alpha}$  is a set of the same type and since the product of two such sets yields a third.

Let  $M$  be any set of real numbers; let  $\{\Delta_{\alpha}\}$  be the set of all coverings of  $M$ . Then  $\prod_{\alpha} E(\Delta_{\alpha})$  is a projection by Theorem 3.4. This projection is called the exterior projection measure of  $M$ , and we write  $\prod_{\alpha} E(\Delta_{\alpha}) = E[M]$ . The set of all projections  $E[M]$  can be embedded in a  $K$ -bounded lattice by Theorem 3.4.

We discuss at this point some matters of future usefulness. If  $M_1 \supset M_2$ , then  $E[M_1] \geq E[M_2]$ . It is also clear that for any interval  $\delta$ ,  $E[\delta] < E(\delta)$ . We shall show that  $E[\delta] = E(\delta)$ . Let  $\Delta = \sum_{\alpha=1}^{\infty} \delta_{\alpha}$  be any covering of  $\delta = \{a < \lambda \leq b\}$ . We may and shall assume that  $\Delta = \delta$ . Let  $\delta_0 = \{r < \lambda \leq a\}$ ,  $\delta_{-1} = \{b < \lambda \leq R\}$ . Then by the discussion preceding this theorem,

$$\sum_{\alpha=1}^{\infty} E(\delta_{\alpha}) = E(\delta_{-1}) \dot{+} E(\delta_0) \dot{+} \sum_{\alpha=1}^{\infty} E(\delta_{\alpha}) = I.$$

Since

$$E(\delta_{-1}) \dot{+} E(\delta_0) \dot{+} E(\delta) = I, \quad E(\delta_i)E(\delta) = E(\delta_i)E(\delta_j) = 0, \\ i = -1, 0; j = 1, 2, \dots,$$

we conclude  $\sum_1^* E(\delta_\alpha) = E(\delta)$ . Hence  $E[\delta] = E(\delta)$ .

Similarly, if  $\Delta = \sum_1^* \delta_\alpha$ , then  $E[\Delta] = E(\Delta)$  ( $= \sum_1^* E(\delta_\alpha)$  by definition). Clearly  $E[\Delta] < E(\Delta)$ . Let  $\sum_1^* \delta'_\alpha$  be a covering of  $\Delta$ ; then  $\sum_1^* \delta'_\alpha \supset \delta_i$ , ( $i = 1, 2, \dots$ ); hence  $\sum_1^* E(\delta'_\alpha) > E[\delta_i] = E(\delta_i)$ . Thus  $\sum_1^* E(\delta'_\alpha) > \sum_1^* E[\delta_\alpha] = \sum_1^* E(\delta_\alpha) = E(\Delta)$ .

For any set  $M$ , and any element  $f \in \mathfrak{B}$ , there exists a sequence of coverings  $\{\Delta'_n\}$  of  $M$  such that  $E(\Delta'_n)f \rightarrow E[M]f$ . For there exist coverings  $\Delta_n$  such that  $\prod_1^n E(\Delta_n)f \rightarrow E[M]f$  (see proof of Theorem 3.3). Now  $\prod_1^n \Delta_n$  is a set of the " $\Delta$  type"; write  $\Delta'_n = \prod_1^n \Delta_n$ . Then we obtain  $\prod_1^n E(\Delta_n) > E[\Delta'_n] = E(\Delta'_n) > E[M]$ . Hence  $E(\Delta'_n)f \rightarrow E[M]f$ .

If  $A_n, B_n$  are projections in a  $K$ -bounded lattice  $\Omega$  such that for a given  $f \in \mathfrak{B}$ ,  $\|(A_n - B_n)f\| \leq \epsilon_n$ , ( $n = 1, 2, \dots$ ), then for any  $C \in \Omega$ ,

$$\left\| C \left( \sum_1^n A_\alpha - \sum_1^n B_\alpha \right) f \right\| \leq K(\epsilon_1 + 2\epsilon_2 + \dots + 2^{n-1}\epsilon_n).$$

For  $n=1$ ,  $\|(A_1 - B_1)f\| \leq \epsilon_1$ ,  $\|C(A_1 - B_1)f\| \leq K\epsilon_1$ . Assume the statement for  $n-1$ . Then

$$C \left( \sum_1^n A_\alpha - \sum_1^n B_\alpha \right) f = C \left( \sum_2^n A_\alpha - \sum_2^n B_\alpha \right) f + C(A_1 - B_1)f \\ - C \left( A_1 \sum_2^n A_\alpha - B_1 \sum_2^n B_\alpha \right) f.$$

Hence

$$\left\| C \left( \sum_1^n A_\alpha - \sum_1^n B_\alpha \right) f \right\| \leq K(\epsilon_2 + \dots + 2^{n-2}\epsilon_n) \\ + \left\| \left[ CA_1 \left( I - \sum_2^n A_\alpha \right) - CB_1 \left( I - \sum_2^n B_\alpha \right) \right] f \right\|.$$

The last term above does not exceed  $K\epsilon_1 + K(\epsilon_2 + \dots + 2^{n-2}\epsilon_n)$ .

**THEOREM 4.2.** For any set  $M$  and its complement  $\overline{M}$ ,  $E[M] \dot{+} E[\overline{M}] = I$ .

For any  $f \in \mathfrak{B}$ , we choose sequences  $\{\Delta_n\}$ ,  $\Delta_n \supset M$ ,  $\{\Delta'_n\}$ ,  $\Delta'_n \supset \overline{M}$ , such that  $E(\Delta_n)f \rightarrow E[M]f$ ,  $E(\Delta'_n)f \rightarrow E[\overline{M}]f$ . By the discussion preceding Theorem 4.1,  $E(\Delta_n)f \dot{+} E(\Delta'_n)f = f$ . But

$$E(\Delta_n)f \dot{+} E(\Delta'_n)f \rightarrow E[M]f \dot{+} E[\overline{M}]f$$

by the last remark preceding this theorem. This establishes our assertion.

THEOREM 4.3. If  $M = \sum_1^\infty M_\alpha$ ,  $E[M] = \sum_1^\infty E[M_\alpha]$ .

Since  $M \supset M_n$ ,  $E[M] \geq E[M_n]$ , ( $n=1, 2, \dots$ ); hence  $E[M] \geq \sum_1^\infty E[M_\alpha]$ . For fixed  $f \in \overline{M}_{E[M]}$  we find coverings  $\Delta_n \supset M_n$ , ( $n=1, 2, \dots$ ), such that  $\|(E(\Delta_n) - E[M_n])f\| \leq \epsilon_n$  where  $\epsilon_n \geq 0$ . Since  $\sum_1^\infty \Delta_n \supset M$ ,  $\sum_1^\infty E(\Delta_n)f = f$ . Thus for  $\epsilon > 0$  we may find an integer  $n$  such that  $\|f - \sum_1^n E(\Delta_n)f\| \leq \epsilon$ . Finally,

$$\begin{aligned} \left\| f - \sum_1^n E[M_\alpha]f \right\| &\leq \left\| f - \sum_1^n E(\Delta_n)f \right\| + \left\| \sum_1^n E(\Delta_n)f - \sum_1^n E[M_\alpha]f \right\| \\ &\leq \epsilon + K(\epsilon_1 + 2\epsilon_2 + \dots + 2^{n-1}\epsilon_n) \leq \epsilon(1 + K) \end{aligned}$$

if  $\epsilon_n < \epsilon/2^{n-1}$ . Thus  $\sum_1^\infty E[M_\alpha]f = f$ , and  $E[M] = \sum_1^\infty E[M_\alpha]$ .

A set  $M$  is said to be projection measurable if  $E[M] \cdot E[\overline{M}] = 0$  where  $\overline{M}$  denotes the complement of  $M$ . If  $M$  is projection measurable, its projection measure is defined to be  $E[M]$ . If  $M$  is projection measurable,  $\overline{M}$  is projection measurable. Since  $E[M] + E[\overline{M}] = I$ ,  $E[M] \cdot E[\overline{M}] = 0$ ,  $E[\overline{M}] = I - E[M]$ . Any set  $\delta$  is projection measurable.

THEOREM 4.4. If  $M_\alpha$  is projection measurable, ( $\alpha=1, 2, \dots$ ), then  $M = \sum_1^\infty M_\alpha$  is projection measurable and  $E[M] = \sum_1^\infty E[M_\alpha]$ . In addition  $M' = \prod_1^\infty M_\alpha$  is projection measurable and  $E[M'] = \prod_1^\infty E[M_\alpha]$ .

We have  $\overline{M} = \prod_1^\infty \overline{M}_\alpha$ , hence  $E[\overline{M}] \leq \prod_1^\infty E[\overline{M}_\alpha]$ . By Theorem 4.3,  $E[M] = \sum_1^\infty E[M_\alpha]$ . We shall prove that  $E[M] \cdot \prod_1^\infty E[\overline{M}_\alpha] = 0$ . This will imply  $E[M] \cdot E[\overline{M}] = 0$ ,  $M$  projection measurable. For fixed  $\epsilon > 0$ ,  $f \in \mathfrak{B}$ , we determine  $n$  so that

$$\left\| E[M]f - \sum_1^n E[M_\alpha]f \right\| \leq \epsilon.$$

Then

$$\begin{aligned} \left\| E[M] \cdot \prod_1^\infty E[\overline{M}_\alpha]f \right\| &\leq \left\| \left( E[M] - \sum_1^n E[M_\alpha] \right) \prod_1^\infty E[\overline{M}_\alpha]f \right\| \\ &\quad + \left\| \left( \sum_1^n E[M_\alpha] \right) \cdot \prod_1^\infty E[\overline{M}_\alpha]f \right\| \leq K\epsilon. \end{aligned}$$

For  $M'$ , we have  $\overline{M}' = \sum_1^\infty \overline{M}_\alpha$ , and  $\overline{M}'$  is projection measurable according to what precedes; hence  $M'$  is projection measurable, and

$$E[\overline{M}'] = \sum_1^\infty E[\overline{M}_\alpha] = \sum_1^\infty (I - E[M_\alpha]).$$

Thus if the manifolds associated to  $E[M_\alpha]$  are  $\mathfrak{M}_\alpha$ ,  $\mathfrak{N}_\alpha$ , those associated to  $\sum_1^\infty (I - E[M_\alpha])$  are  $\sum_1^\infty \mathfrak{N}_\alpha$ ,  $\prod_1^\infty \mathfrak{M}_\alpha$ , and those associated to  $E[M'] = I - E[\overline{M}']$  are  $\prod_1^\infty \mathfrak{N}_\alpha$ ,  $\sum_1^\infty \mathfrak{M}_\alpha$ . In other words,  $E[M'] = \prod_1^\infty E[M_\alpha]$ .

From this point forward, and for obvious reasons, we shall denote the projection measure of a projection measurable set  $M$  by  $E(M)$ . Thus the new symbol  $E(M)$  is identical with the old whenever both are significant. We note that Borel sets are projection measurable. We shall use the terms "of zero measure," "almost everywhere," and so on, without introducing them formally.

## V. THE OPERATIONAL CALCULUS

In this chapter we establish the existence of an extensive homomorphism between a substantial class of real functions and a class of operators. As our principal theorem indicates, the correspondence is more than a ring homomorphism.

To a resolution of the identity  $E(\lambda)$ , we associate an operator  $A$  in the following manner: Let  $\delta_1, \dots, \delta_n$  be nonoverlapping intervals covering the fundamental interval  $r < \lambda \leq R$  (where  $E(r) = 0$ ,  $E(R) = I$ ). Let  $\lambda_i \in \delta_i$ , ( $i = 1, \dots, n$ ), and form the operator  $\sum_1^n \lambda_i E(\delta_i)$ . Consider, as in the classic case, a sequence of such divisions of the fundamental interval in which the maximum length of any interval converges to zero. The sequence of operators which corresponds to the sequence of subdivisions converges by virtue of Definition 4, (3) to a linear operator  $A$  whose bound does not exceed  $K \cdot \max(|r|, |R|)$ . Because of its suggestive value, we may, if we wish, write  $A = \int \lambda dE(\lambda)$ .

We now apply our theory of measure to real functions  $\phi(\lambda)$  of the real variable  $\lambda$ . The function  $\phi(\lambda)$  is said to be  $E(\lambda)$ -measurable if the sets  $M_\mu = \{\phi(\lambda) \leq \mu\}$  are projection measurable. We consider exclusively functions  $\phi(\lambda)$  which are measurable and bounded almost everywhere. If  $\phi_1(\lambda)$  and  $\phi_2(\lambda)$  are two such functions, so are  $\phi_1(\lambda) + \phi_2(\lambda)$ ,  $\phi_1(\lambda) \cdot \phi_2(\lambda)$ . Any Baire function is  $E(\lambda)$ -measurable. The limit of a converging sequence of  $E(\lambda)$ -measurable functions is  $E(\lambda)$ -measurable.

**THEOREM 5.1.** *Let  $\phi(\lambda)$  be  $E(\lambda)$ -measurable and bounded almost everywhere. Let  $M_\mu = \{\phi(\lambda) \leq \mu\}$ , ( $-\infty < \mu < \infty$ ). Then  $D(\mu) = E(M_\mu)$  is a resolution of the identity.*

Since  $\phi(\lambda)$  is bounded,  $|\phi(\lambda)| < s$  almost everywhere. To establish (1) in Definition 4, note that  $D(-s) = E(M_{-s}) = 0$ ,  $D(s) = E(M_s) = I$ . (2) Since  $M_\mu \subset M_\nu$  for  $\mu < \nu$ , we have  $D(\mu) \subset D(\nu)$ . (3) Let  $\delta_1, \dots, \delta_n$  be nonoverlapping intervals in the  $\mu$ -space and  $\mu_1, \dots, \mu_n$  complex numbers for which  $|\mu_i| \leq 1$ . Then to the  $\delta_i$  correspond measurable sets  $M_i$  in the  $\lambda$ -space for which  $M_i M_j = 0$ , ( $i \neq j$ ). For fixed  $f \in \mathfrak{B}$  and  $\epsilon > 0$  we find  $\Delta_i = \sum_1^n \delta_{i\alpha}$ , ( $i = 1, \dots, n$ ), such that

$$\left\| \left[ \sum_1^{\infty} {}^*E(\delta_{ia}) - D(\delta_i) \right] f \right\| < \epsilon$$

(we assume that  $\delta_{ij}\delta_{ik}=0$ , ( $j \neq k$ )). We then find suitable integers  $n_i$  such that

$$\left\| \left[ \sum_1^{n_i} {}^*E(\delta_{ia}) - D(\delta_i) \right] f \right\| < \epsilon.$$

Let us replace  $\sum_1^{n_i} {}^*E(\delta_{ia})$  by  $E_i$ ,  $D(\delta_i)$  by  $D_i$ . Then

$$\begin{aligned} \sum_1^n \mu_a D_a &= \sum_1^n \mu_a (D_a - E_a) \\ &+ \left[ \mu_1 E_1 + \mu_2 (E_2 - E_1 E_2) + \cdots + \mu_n \left( E_n - E_n \sum_1^{n-1} {}^*E_a \right) \right] \\ &+ \mu_2 E_2 E_1 + \cdots + \mu_n E_n \sum_1^{n-1} {}^*E_a. \end{aligned}$$

From the terms on the right side of the equation we obtain  $\|\sum_1^n \mu_a (D_a - E_a) f\| \leq n\epsilon$ . The operator in brackets yields, when applied to  $f$ , an element of norm at most  $K\|f\|$  by Definition 4, (3). In examining the norm of the element  $E_i \sum_1^{i-1} {}^*E_a f$ , ( $i=2, \dots, n$ ), we use the fact that  $E(M_i) \sum_1^{i-1} {}^*E(M_a) = 0$  as well as an inequality immediately preceding Theorem 4.2. We have

$$\begin{aligned} \left\| E_i \left( \sum_1^{i-1} {}^*E_a \right) f \right\| &\leq \left\| E_i \left( \sum_1^{i-1} {}^*E_a - \sum_1^{i-1} {}^*E(M_a) \right) f \right\| \\ &+ \left\| \sum_1^{i-1} {}^*E(M_a) (E_i - E(M_i)) f \right\| \\ &\leq K\epsilon(1 + 2 + \cdots + 2^{i-2}) + K\epsilon \\ &= 2^{i-1} K\epsilon. \end{aligned}$$

Thus  $\|(\sum_1^n \mu_a D_a) f\| \leq n\epsilon + K\|f\| = 2(2^{n-1} - 1)K\epsilon$ . This proves (3) since  $n$  is fixed at the outset.

As  $E(\lambda)$  yields an operator  $A$ ,  $A = \int \lambda dE(\lambda)$ , so does  $D(\mu)$  generate an operator which we designate by  $\phi(A)$ ,  $\phi(A) = \int \mu dD(\mu)$ . We have thus established a correspondence between  $E(\lambda)$ -measurable functions  $\phi(\lambda)$  and operators  $\phi(A)$ . We write this correspondence in the form  $\phi(\lambda) \sim \phi(A)$ . In particular, we have  $\lambda \sim A$ . In what follows, the statements  $\phi(\lambda) = \psi(\lambda)$ ,  $\phi(\lambda) < \psi(\lambda)$ ,  $\phi_n(\lambda) \rightarrow \phi(\lambda)$ ,  $\dots$  will imply equality, inequality, convergence, and so on, respectively, almost everywhere.

**THEOREM 5.2.** *Let  $E(\lambda)$  be a resolution of the identity, and let  $A = \int \lambda dE(\lambda)$ . Let  $\phi(\lambda)$ ,  $\psi(\lambda)$ ,  $\phi_n(\lambda)$  be any bounded  $E(\lambda)$ -measurable functions. The corre-*

*spondence*  $\sim$  which associates to these functions the operators  $\phi(A), \psi(A), \phi_n(A)$  has the following properties:

- (1) If  $|\phi(\lambda)| \leq s$ , then  $|\phi(A)| \leq Ks$ .
- (2)  $\phi(\lambda) + \psi(\lambda) \sim \phi(A) + \psi(A)$ .
- (3)  $\phi(\lambda) \cdot \psi(\lambda) \sim \phi(A) \cdot \psi(A)$ .
- (4) If  $|\phi_n(\lambda)| < b$  and if  $\{\phi_n(\lambda)\}$  converges to  $\phi(\lambda)$ , then  $\{\phi_n(A)\}$  converges to  $\phi(A)$ .
- (5)  $\phi(A)$  is permutable with any linear operator with which  $A$  is permutable.
- (6) If  $D(\mu)$  is the resolution of the identity of  $B = \phi(A)$  and if  $\xi(\mu)$  is  $D(\mu)$ -measurable, then  $\xi(\phi(\lambda))$  is  $E(\lambda)$ -measurable and  $\xi(\phi(\lambda)) \sim \xi(B)$ .
- (7)  $\phi(A) = 0$  if and only if  $\phi(\lambda) = 0$ .
- (8)  $\phi(A)$  is a projection if and only if  $\phi(\lambda)$  assumes only the values 0 and 1.

If  $\phi(\lambda)$  is a function assuming the values  $\lambda_1, \dots, \lambda_n$  only on the sets  $M_1, \dots, M_n$ , respectively, then  $\phi(A) = \sum_1^n \lambda_n E(M_n)$ . If  $\phi(\lambda)$  is arbitrary, we may approximate uniformly to  $\phi(\lambda)$  by functions  $\phi_n(\lambda)$  assuming only a finite number of values.  $\phi_n(\lambda)$  may, for instance, be defined as follows: On  $\{q/n < \phi(\lambda) \leq (q+1)/n\}$ , ( $q=0, \pm 1, \pm 2, \dots$ ),  $\phi_n(\lambda) = q/n$  for  $q=0, 1, 2, \dots$ , and  $\phi_n(\lambda) = (q+1)/n$  for  $q=-1, -2, \dots$ . By the definition of  $\phi(A)$ ,  $\phi_n(A) \rightarrow \phi(A)$ . We note that if  $|\phi(\lambda)| \leq s$ ,  $|\phi_n(\lambda)| \leq s$ , by Theorem 5.1,  $|\phi_n(A)| \leq Ks$ , ( $n=1, 2, \dots$ ); hence  $|\phi(A)| \leq Ks$ . This proves (1).

(2) Given  $\phi(\lambda)$  and  $\psi(\lambda)$ , let  $\{\phi_n(\lambda)\}$  and  $\{\psi_n(\lambda)\}$  be chosen as indicated in the previous paragraph. Then  $\phi_n(\lambda) + \psi_n(\lambda) \rightarrow \phi(\lambda) + \psi(\lambda)$  uniformly. The functions  $\phi_n(\lambda) + \psi_n(\lambda)$ , ( $n=1, 2, \dots$ ), assume only a finite number of values, and clearly  $\phi_n(\lambda) + \psi_n(\lambda) \sim \phi_n(A) + \psi_n(A) = B_n$ . If we write  $\phi(\lambda) + \psi(\lambda) \sim C$ , then by the first paragraph,  $C$  is the limit of a sequence of operators  $C_n$  which has the property that  $|B_n - C_n| \rightarrow 0$ . Hence  $B_n \rightarrow C$  or  $C = \phi(A) + \psi(A)$ .

(3) The relation is derived by replacing, in the previous paragraph,  $\phi + \psi$  by  $\phi \cdot \psi$ ,  $\phi_n + \psi_n$  by  $\phi_n \cdot \psi_n$ .

(4) First, let  $\{\phi_n(\lambda)\}$  be a monotone decreasing sequence of positive functions,  $b \geq \phi_1(\lambda) \geq \phi_2(\lambda) \geq \dots$ , for which  $\phi_n(\lambda) \rightarrow 0$ . Let  $f \in \mathfrak{B}$  and  $\epsilon > 0$ . Let  $M_n = \{\phi_n(\lambda) > \epsilon\}$ . Then  $M_1 \supset M_2 \supset \dots$ , and, since  $\phi_n(\lambda) \rightarrow 0$ ,  $\prod_1^\infty M_n = 0$ ; hence  $\prod_1^\infty E(M_n) = 0$ . Thus there exists an integer  $r$  such that

$$\left\| \prod_1^r E(M_n) f \right\| = \|E(M_r) f\| \leq \epsilon.$$

Now

$$\begin{aligned} \|\phi_{r+s}(A) f\| &\leq \|\phi_{r+s}(A)(I - E(M_r)) f\| + \|\phi_{r+s}(A) E(M_r) f\| \\ &\leq \epsilon K(\|f\| + b), \end{aligned} \quad s = 1, 2, \dots,$$

by applying to the first term on the right of the inequality the results (3)

and (1). Thus  $\phi_n(A) \rightarrow 0$ . Similarly if  $-b \leq \phi_1(\lambda) \leq \phi_2(\lambda) \leq \dots$  and  $\phi_n(\lambda) \rightarrow 0$ , then  $\phi_n(A) \rightarrow 0$ .

Now let  $\{\phi_n(\lambda)\}$ ,  $|\phi_n(\lambda)| \leq b$ , be arbitrary subject to the restriction  $\phi_n(\lambda) \rightarrow 0$ . Let  $\bar{\psi}_n(\lambda) = \max \{\phi_n(\lambda), \phi_{n+1}(\lambda), \dots\}$ ,  $\underline{\psi}_n(\lambda) = \min \{\phi_n(\lambda), \phi_{n+1}(\lambda), \dots\}$ . Then  $\underline{\psi}_n(\lambda) \leq \phi_n(\lambda) \leq \bar{\psi}_n(\lambda)$ ,  $\bar{\psi}_n(\lambda) \rightarrow 0$ ,  $\underline{\psi}_n(\lambda) \rightarrow 0$ ,  $\bar{\psi}_1(\lambda) \geq \bar{\psi}_2(\lambda) \geq \dots$ , and  $\underline{\psi}_1(\lambda) \leq \underline{\psi}_2(\lambda) \leq \dots$ . Thus for fixed  $f \in \mathfrak{B}$  and  $\epsilon > 0$ , we may find an integer  $r$  and sets  $\bar{M}_r, \underline{M}_r$  such that  $\|E(\bar{M}_r)f\| \leq \epsilon$ ,  $\|E(\underline{M}_r)f\| < \epsilon$ ,  $\bar{\psi}_r(\lambda) \leq \epsilon$  except on  $\bar{M}_r$ ,  $\underline{\psi}_r(\lambda) \geq -\epsilon$  except on  $\underline{M}_r$ . Then

$$\begin{aligned} \|\phi_{r+s}(A)f\| &\leq \|\phi_{r+s}(A)(I - [E(\bar{M}_r) + E(\underline{M}_r)])f\| + \|\phi_{r+s}(A)[E(\bar{M}_r) + E(\underline{M}_r)]f\| \\ &\leq \epsilon K(\|f\| + 2b), \end{aligned} \quad s = 1, 2, \dots$$

Thus again  $\phi_n(A) \rightarrow 0$ .

In the general case,  $\phi_n(\lambda) \rightarrow \phi(\lambda)$ ; hence  $\psi_n(\lambda) = \phi_n(\lambda) - \phi(\lambda) \rightarrow 0$ . By (2) and above,  $\psi_n(A) = \phi_n(A) - \phi(A) \rightarrow 0$ ,  $\phi_n(A) \rightarrow \phi(A)$ . This completes the proof of statement (4).

(5) In the first place, a linear operator  $B$  is permutable with  $A$ ,  $BA = AB$ , if and only if  $BE(\lambda) = E(\lambda)B$ ,  $(-\infty < \lambda < \infty)$ . If the latter equation holds, so does the former from the very definition of  $A$ . We prove the converse. For any  $\mu$ , let  $\{\phi_{\mu n}(\lambda)\}$  be a sequence of polynomials such that  $|\phi_{\mu n}(\lambda)| \leq C$  in the interval  $r < \lambda \leq R$ , and  $\phi_{\mu n}(\lambda) \rightarrow \phi_\mu(\lambda)$  where  $\phi_\mu(\lambda) = 1$ ,  $(\lambda \leq \mu)$ ,  $\phi_\mu(\lambda) = 0$ ,  $(\lambda > \mu)$ . We have  $\phi_\mu(A) = E(\mu)$ ,  $\phi_{\mu n}(A) \rightarrow E(\mu)$  by (4), and  $\phi_{\mu n}(A)B = B\phi_{\mu n}(A)$  by (2) and (3). Hence  $E(\mu)B = BE(\mu)$ .

If  $D(\mu)$  is the resolution of the identity of  $\phi(A)$ , then  $D(\mu)$ , and hence  $\phi(A)$ , are permutable with any linear operator permutable with  $E(\lambda)$  (Theorem 3.3 and others), and hence, by the above, permutable with any linear operator permutable with  $A$ .

(6) Let  $\phi(\lambda)$  be  $E(\lambda)$ -measurable, and let  $D(\mu)$  be the resolution of the identity of  $\phi(A)$ . Let  $M$  be any  $D(\mu)$ -measurable set, and let  $N$  denote the set of all numbers  $\lambda$  such that  $\phi(\lambda) \in M$ . We shall show that  $N$  is  $E(\lambda)$ -measurable and that  $D(M) = E(N)$ . As before, we denote the exterior measure of a set  $H$  by  $D[H]$  (or  $E[H]$ ); use of the symbol  $D(H)$  (or  $E(H)$ ) will imply that  $H$  is measurable. Assume  $f \in \mathfrak{B}$  and  $\epsilon > 0$ . Then there exists a set  $\Delta = \Delta(\epsilon)$  covering  $M$  such that  $\|D(\Delta)f - D(M)f\| < \epsilon$ . Let  $\Gamma = \Gamma(\epsilon)$  denote the set of all  $\lambda$  such that  $\phi(\lambda) \in \Delta$ . Then  $\Gamma$  is  $E(\lambda)$ -measurable,  $D(\Delta) = E(\Gamma)$ , and  $\Gamma \supset N$ . Let  $\{\epsilon_n\}$  be a sequence converging to 0 with  $1/n$ , and let  $\Theta_f = \prod_1^\infty \Delta(\epsilon_n)$ ,  $\Psi_f = \prod_1^\infty \Gamma(\epsilon_n)$ . Then  $\Theta_f \supset M$ ,  $\Psi_f \supset N$ ,  $D(\Theta_f)f = D(M)f = E(\Psi_f)f$ . Let  $\Theta = \prod_f \Theta_f$ ,  $\Psi = \prod_f \Psi_f$ , where products are taken over the entire space  $\mathfrak{B}$ . Then  $\Theta_f \supset \Theta \supset M$ ,  $\Psi_f \supset \Psi \supset N$ . Therefore  $\prod_f D(\Theta_f) > D[\Theta] > D(M)$ . But we see readily that  $\prod_f D(\Theta_f) = D(M)$ . Hence  $D(M) = \prod_f D(\Theta_f) = \prod_f E(\Psi_f) > E[\Psi] > E[N]$ .

If we now turn our attention to  $\overline{M}$ , the set of all  $\lambda$  such that  $\phi(\lambda) \in \overline{M}$  is precisely  $\overline{N}$ . The argument just given leads to the conclusion  $D(\overline{M}) > E[\overline{N}]$ . Since  $M$  is measurable,  $D(M) \cdot D(\overline{M}) = 0$ ; hence  $N$  is measurable and  $E(N) = D(M)$ .

If  $\xi(\mu)$  is  $D(\mu)$ -measurable, the set  $M_\nu = \{\xi(\mu) \leq \nu\}$  is  $D(\mu)$ -measurable, the set  $N_\nu = \{\xi(\phi(\lambda)) \leq \nu\}$  is  $E(\lambda)$ -measurable and  $E(N_\nu) = D(M_\nu)$ . Thus  $\xi(\phi(\lambda))$  is  $E(\lambda)$ -measurable. The operator  $\xi(B)$ , where  $B = \phi(A)$ , has the resolution of the identity  $J(\nu) = D(M_\nu)$ . The operator corresponding to the function  $\xi(\phi(\lambda))$  has the resolution of the identity  $J'(\nu) = E(N_\nu)$ . Since  $J(\nu) = J'(\nu)$ , the operators are identical or  $\xi(\phi(\lambda)) \sim \xi(B)$ .

(7) Suppose  $\phi(\lambda) \neq 0$ ; then clearly  $\phi(A) \neq 0$ . Now assume that  $\phi(\lambda) \neq 0$ ; in this case we may assume that  $M = \{|\phi(\lambda)| \geq 1\}$  has an  $E(\lambda)$  measure different from zero, for any other case quickly reduces to this one. Let  $\mathfrak{P}(\lambda)$  be the function defined by  $\mathfrak{P}(\lambda) = 0$  for  $|\lambda| < 1$  and  $\mathfrak{P}(\lambda) = 1$  for  $\lambda \geq 1$ , and let  $\mathfrak{P}_n(\lambda)$  be polynomials such that  $\mathfrak{P}_n(\lambda) \rightarrow \mathfrak{P}(\lambda)$ ,  $|\mathfrak{P}_n(\lambda)| \leq C$  on the interval  $r < \lambda \leq R$ , and  $\mathfrak{P}_n(0) = 0$ . Then  $\mathfrak{P}_n(\phi(\lambda)) \rightarrow \mathfrak{P}(\phi(\lambda))$ , and if we write  $B = \phi(A)$ , we deduce, using (6) and (4), that  $\mathfrak{P}_n(B) = \mathfrak{P}_n(\phi(A)) \rightarrow E(M) \neq 0$ . If  $B = 0$ ,  $\mathfrak{P}_n(B) = 0$  by (2) and (3); hence  $B = \phi(A) \neq 0$ .

(8) If  $\phi(\lambda)$  assumes the values 0 and 1 only, then if  $\phi(\lambda) \sim R$ , since  $(\phi(\lambda))^2 = \phi(\lambda)$ ,  $R^2 = R$  by (3) and  $R$  is a projection. If  $R \sim \phi(\lambda)$  and  $R$  is a projection, then by (2) and (3),  $0 = R^2 - R \sim (\phi(\lambda))^2 - \phi(\lambda) = \psi(\lambda)$ . By (7)  $\psi(\lambda) = 0$ ; hence  $\phi(\lambda)$  assumes the values 0 and 1 only.

As the aim of this presentation has been to establish the possibility of developing an operational calculus in reflexive spaces, we have purposely refrained from doing this in its most general form. Obvious generalizations of our results will present themselves to the reader; these offer, for the most part, no difficulties.

BARNARD COLLEGE, COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

## THERMAL STRESSES IN ELASTIC PLATES\*

BY

I. S. SOKOLNIKOFF AND E. S. SOKOLNIKOFF

1. **Introduction.** Thermal changes in an elastic body are accompanied by shifts in the relative positions of the particles composing the body. Such shifts, in general, cannot proceed freely, and thermal stresses are set up in the body. The analytical basis for the determination of such stresses was provided by Duhamel† and Neumann‡ who, starting from certain assumptions, have modified the stress-strain relations of Hooke. The theory based on the law formulated by Duhamel has not been very much developed§ because of the complicated character of the partial differential equations satisfied by the stress components. In particular, there is a notable lack of a careful formulation of the differential equations governing the deflection of elastic plates subjected to nonuniform distribution of temperatures.

Starting with the usual assumptions of the thin plate theory, Nádai|| has developed the differential equation for deflection of a thin elastic plate subjected to a linear distribution of temperature in the direction of the thickness of the plate. In two recent papers Marguerre¶ has considered some related problems, and, on the basis of reasoning essentially similar to that of Nádai, was led to a somewhat more general equation. In the present paper the differential equation governing the deflection of an elastic plate is derived without using the objectionable assumption of the thin plate theory. It will be seen that the differential equations obtained by Nádai and Marguerre are special cases of the more general equation here given. Furthermore, the hypotheses of Nádai and Marguerre that a heated plate will be in a state of plane or generalized plane stress (which form a focal point in their discussion) are not used in this paper. The abrogation of these hypotheses enormously complicates the analysis, but inasmuch as there has been some considerable doubt regarding the validity of the generalized stress assumption, it appears necessary to sacrifice simplicity. The complexity of the situation lies in the nature of the problem itself.

\* Presented to the Society, April 8, 1938; received by the editors February 9, 1938.

† J. M. C. Duhamel, *Mémoires . . . par Divers Savants*, vol. 5, 1838, p. 440.

‡ F. E. Neumann, *Vorlesungen über die Theorie der Elasticität der festen Körper*, 1885.

§ Cf. A. E. H. Love, *Mathematical Theory of Elasticity*, 4th edition, 1927, p. 109.

|| A. Nádai, *Elastische Platten*, 1925, p. 268.

¶ K. Marguerre, *Zeitschrift für angewandte Mathematik und Mechanik*, vol. 15 (1935), pp. 369-372; *Ingenieur-Archiv*, vol. 8 (1937), pp. 216-228.

2. **General thermo-elastic equations.** According to Duhamel's law the stresses and strains in an elastic body are connected in the following way:

$$(2.1) \quad e_{ij} = \frac{1}{2} \left( \frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) = \frac{1+\sigma}{E} X_{ij} - \left( \frac{\sigma}{E} \Theta + \alpha T \right) \delta_{ij}, \quad i, j = 1, 2, 3,$$

where  $\delta_{ij}$  is the Kronecker delta,  $\Theta = \sum_{i=1}^3 X_{ii}$ ,  $T$  is the prescribed temperature at any point of the body, and  $\alpha$  is a constant depending on the physical properties of the material. Young's modulus  $E$ , and Poisson's ratio  $\sigma$  are regarded as independent of the temperature.\*

Since the equations of equilibrium are deduced with no reference to the law connecting stresses and strains, they remain valid in this case.† They are

$$(2.2) \quad \sum_{j=1}^3 \frac{\partial X_{ij}}{\partial x_j} = 0, \quad i = 1, 2, 3,$$

where the stress components  $X_{ij}$  must satisfy, on the boundary of the solid, the following conditions:

$$(2.3) \quad \sum_{j=1}^3 X_{ij} \cos(x_j, n) = X_{ni}, \quad i = 1, 2, 3.$$

It is well known that the satisfaction of (2.2) and (2.3) does not guarantee a physically realizable system of stresses. The additional conditions which form a connecting link between (2.1) and (2.2) are the compatibility equations of St. Venant‡ which demand, in effect, that the displacements  $u_i$  in a simply connected region be single-valued functions. Substitution of (2.1) in St. Venant's compatibility equations gives, after some reduction, the desired equations of connection:

$$(2.4) \quad \nabla^2 X_{ij} + \frac{1}{1+\sigma} \frac{\partial^2 \Theta}{\partial x_i \partial x_j} = - \frac{\alpha E}{1-\sigma} \nabla^2 T \delta_{ij} - \frac{\alpha E}{1+\sigma} \frac{\partial^2 T}{\partial x_i \partial x_j},$$

$i, j = 1, 2, 3,$

where  $\Theta$  satisfies the equation

$$(2.5) \quad \nabla^2 \Theta = - \frac{2\alpha E}{1-\sigma} \nabla^2 T$$

and

\* For some critical remarks regarding the applicability of these equations in practice, see a paper by J. N. Goodier, *Philosophical Magazine*, vol. 23 (1937), pp. 1017-1032.

† A. E. H. Love, loc. cit., p. 100.

‡ A. E. H. Love, loc. cit., p. 49, equations (25).

$$\nabla^2 \equiv \sum_{i=1}^3 \frac{\partial^2}{\partial x_i \partial x_i}.$$

The equations (2.4) reduce, as they should, to the compatibility equations of Beltrami\* when  $T$  is set equal to a constant. The sets of equations (2.2), (2.3), and (2.4) give a unique determination of stresses.

In order to avoid the use of the subscripts on the variables  $u_3$  and  $x_3$ , which figure prominently in the remainder of this paper, the letters  $w$  and  $z$  will be used in their stead. The notation adopted at this point agrees with that of Love's treatise.

One can readily obtain a set of equations in the displacements  $u$ ,  $v$ , and  $w$  by substituting the expressions for the stress components from (2.1) in (2.2). The resulting equations† are

$$(2.6) \quad \nabla^2 w = \frac{2\alpha(1+\sigma)}{1-2\sigma} \frac{\partial T}{\partial z} - \frac{1}{1-2\sigma} \frac{\partial \Delta}{\partial z},$$

and two similar equations for  $u$  and  $v$ , where

$$\Delta \equiv \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}.$$

It will be observed, with reference to (2.1), that

$$(2.7) \quad \Delta = \frac{1-2\sigma}{E} \Theta + 3\alpha T.$$

The foregoing equations give a unique characterization of the behavior of a simply connected elastic body subjected to heat, and involve no simplifying assumptions in regard to the geometrical properties of the body. In deriving the thermo-elastic plate equation, both Nádai and Marguerre assume that the plate is so thin that one is permitted to write

$$(2.8) \quad u = -z \frac{\partial w_0}{\partial x}, \quad v = -z \frac{\partial w_0}{\partial y},$$

where  $w_0$  denotes the deflection of the "middle surface" of the plate, and is a function of  $x$  and  $y$  only. They further assume that the plate is in a state of plane stress, so that one is permitted to set

$$(2.9) \quad Z_z = 0$$

throughout the thickness of the plate.

\* A. E. H. Love, loc. cit., p. 135.

† Cf. S. Timoshenko, *Theory of Elasticity*, 1934, p. 205.

A number of investigators in the theory of elasticity have objected to the simplifying assumption (2.9), and have attempted to justify it on the basis of something more satisfactory than intuitive feeling or experimental grounds. The latest of such attempts is that of Southwell\* who investigates the distribution of stress along the edges of the plate where it is assumed that the plate is in a state of generalized plane stress. His results, although not quite conclusive, point to the fact that in order to maintain a state of generalized plane stress one is obliged to apply a complicated distribution of stresses on the edges of the plate of a type which is not likely to be realized in practice. The derivation of the differential equation for the deflection of the middle surface of an elastic plate given in the next section makes no use of the simplifying assumptions (2.8) and (2.9), and thus appears to be applicable to thick as well as thin plates.

**3. Thermo-elastic plate equation.** The two-dimensional Laplacian operator  $\partial^2/\partial x^2 + \partial^2/\partial y^2$  will be denoted here by the symbol  $\nabla_1^2$ , so that

$$\nabla_1^2 w = \nabla^2 w - \frac{\partial^2 w}{\partial z^2}.$$

Substituting (2.7) in (2.6) gives

$$(3.1) \quad \nabla_1^2 w = -\alpha \frac{\partial T}{\partial z} - \frac{1}{E} \frac{\partial \Theta}{\partial z} - \frac{\partial^2 w}{\partial z^2}.$$

Differentiating the stress-strain relations (2.1) for  $e_{33}$  with respect to  $z$  and substituting the resulting value of  $\partial^2 w/\partial z^2$  in (3.1) gives

$$(3.2) \quad \nabla_1^2 w = -2\alpha \frac{\partial T}{\partial z} + \frac{\sigma - 1}{E} \frac{\partial \Theta}{\partial z} - \frac{\sigma + 1}{E} \frac{\partial Z_z}{\partial z}.$$

But the equations (2.4) with  $i=j=3$  give

$$(3.3) \quad \frac{1}{1 + \sigma} \frac{\partial^2 \Theta}{\partial z^2} = -\frac{\alpha E}{1 - \sigma} \nabla^2 T - \frac{\alpha E}{1 + \sigma} \frac{\partial^2 T}{\partial z^2} - \nabla^2 Z_z,$$

and from (2.5),

$$(3.4) \quad \begin{aligned} \nabla_1^2 \Theta &= -\frac{2\alpha E}{1 - \sigma} \nabla^2 T - \frac{\partial^2 \Theta}{\partial z^2} \\ &= -\alpha E \nabla_1^2 T + (1 + \sigma) \nabla^2 Z_z, \end{aligned}$$

where the last step results from the substitution of the value of  $\partial^2 \Theta/\partial z^2$  from (3.3).

\* R. V. Southwell, Philosophical Magazine, vol. 21 (1936), pp. 201-215.

Operating on (3.2) with  $\nabla_1^2$  gives

$$(3.5) \quad \nabla_1^4 w = -2\alpha \frac{\partial}{\partial z} \nabla_1^2 T + \frac{\sigma-1}{E} \frac{\partial}{\partial z} \nabla_1^2 \Theta - \frac{\sigma+1}{E} \frac{\partial}{\partial z} \nabla_1^2 Z_z,$$

and substituting in (3.5) and (3.4) and simplifying, we obtain

$$(3.6) \quad \nabla_1^4 w = -\alpha(1+\sigma)\nabla_1^2 \frac{\partial T}{\partial z} + \frac{(1+\sigma)(\sigma-2)}{E} \frac{\partial}{\partial z} \nabla_1^2 Z_z + \frac{\sigma^2-1}{E} \frac{\partial^3 Z_z}{\partial z^3}.$$

The equation of the middle plane is obtained from (3.6) by setting  $z=0$ . Thus one can write

$$(3.7) \quad \begin{aligned} \nabla_1^4 w_0 = & -\alpha(1+\sigma)\nabla_1^2 \left( \frac{\partial T}{\partial z} \right)_0 + \frac{(1+\sigma)(\sigma-2)}{E} \nabla_1^2 \left( \frac{\partial Z_z}{\partial z} \right)_0 \\ & + \frac{\sigma^2-1}{E} \left( \frac{\partial^3 Z_z}{\partial z^3} \right)_0, \end{aligned}$$

where the zero subscripts denote that the values of the expressions affected are calculated by setting  $z=0$ .

It is interesting to note that if one assumes with Nádai\*

$$T = T_0(x, y) + zT_1(x, y)$$

and  $Z_z=0$ , the equation (3.7) reduces directly to that obtained by Nádai without invoking his additional assumption (2.8).

Equation (3.7) contains in the right-hand member the unknown function  $Z_z$ , the determination of which is given in the following section. It will be seen that in the case of thin plates, for a suitably restricted  $T$ ,  $Z_z$  is quite small. This affords some justification for the assumption (2.9) of earlier investigators.

**4. The determination of  $Z_z$ .** The differential equation satisfied by  $Z_z$  is obtained by setting  $i=j=3$  in (2.4), operating with the Laplacian operator and noting from (2.5) that

$$\nabla^2 \frac{\partial^2 \Theta}{\partial z^2} = -\frac{2\alpha E}{1-\sigma} \nabla^2 \frac{\partial^2 T}{\partial z^2}.$$

The result is

$$(4.1) \quad \nabla^4 Z_z = \frac{\alpha E}{1-\sigma} \left( \frac{\partial^2}{\partial z^2} \nabla^2 T - \nabla^4 T \right).$$

\* Nádai, loc. cit., p. 265.

Denote the thickness of the plate by  $2h$ , and let its faces be given by  $z = \pm h$ ; then it follows from (2.3) that

$$Z_z = Z_n, \quad X_z = X_n, \quad Y_z = Y_n, \quad \text{on } z = \pm h.$$

If the faces of the plate are free from external loads,\* so that the thermal stresses are the only ones under consideration, it is clear that

$$(4.2) \quad \begin{aligned} Z_z(x, y, \pm h) &= 0, \\ Z_x(x, y, \pm h) &= 0, \\ Z_y(x, y, \pm h) &= 0. \end{aligned}$$

From (4.2) and the equations (2.2) it follows that

$$(4.3) \quad \frac{\partial Z_z}{\partial z} = 0, \quad \text{on } z = \pm h.$$

Thus in addition to satisfying (4.1),  $Z_z$  together with  $\partial Z_z / \partial z$  must vanish on the faces of the plate.

If the plate is so large that one is justified in regarding the problem as two-dimensional, the equation (4.1) together with the boundary conditions (4.2) and (4.3) determine  $Z_z$  uniquely. In particular, if the right-hand member of (4.1) is zero (which is certainly the case when  $T$  is steady) and if the plate is infinite, the only solution of the problem is  $Z_z = 0$ . This suggests that in the case of finite plates of thickness small compared with the linear dimensions in the  $x$  and  $y$  directions,  $Z_z$  cannot be very great. In fact, let  $Z_z$  be expanded in an infinite series in powers of  $z$  so that

$$Z_z = \sum_{j=0}^{\infty} a_j(x, y) z^j,$$

where the coefficients  $a_j$  are unknown functions of  $x$  and  $y$ . It will be seen presently that the determination of these unknown functions can be made to depend on the solution of an infinite system of partial differential equations of a complicated sort. In order to make the problem tractable, it is desirable to introduce at this point a simplifying hypothesis† of quite mild nature. It will be assumed that  $Z_z$  is adequately represented by a polynomial in  $z$  of the form

$$(4.4) \quad Z_z = \sum_{j=0}^n a_j(x, y) z^j,$$

\* If the plate is subjected to a normal pressure  $p$ , the first of the boundary conditions reads  $Z_z(x, y, h) = -p$ ,  $Z_z(x, y, -h) = 0$ . The introduction of  $p$  produces no essential complications in the argument that is to follow.

† See in this connection §8 below.

where  $n$  is finite, but can be chosen arbitrarily large. This amounts to assuming that, for sufficiently large  $n$ , the remainder in the Taylor expansion for  $Z_s$  is negligibly small. In order to make the argument clear and not to complicate unduly the resulting equations, the discussion will be confined to a distribution of the temperature of the form\*

$$(4.5) \quad T = \sum_{k=0}^3 T_k(x, y)z^k,$$

which includes the important linear case of Nádai,<sup>†</sup> as well as the cases considered by Timoshenko.<sup>‡</sup> The extension to a more complicated distribution of temperature is obvious and involves little more than the change of the indices of summation.

Noting the operator identity

$$\nabla^4 \equiv \nabla_1^4 + 2 \frac{\partial^2}{\partial z^2} \nabla_1^2 + \frac{\partial^4}{\partial z^4},$$

where

$$\nabla_1^4 \equiv \nabla_1^2 \nabla_1^2,$$

and calculating with its aid  $\nabla^4 Z_s$  defined by (4.4) gives

$$\nabla^4 Z_s = \sum_{j=0}^n [z^j \nabla_1^4 a_j + 2j(j-1)z^{j-2} \nabla_1^2 a_j + j(j-1)(j-2)(j-3)a_j z^{j-4}].$$

The right-hand member of (4.1) upon substituting for  $T$  from (4.5) becomes

$$\frac{\alpha E}{1-\sigma} \left( \frac{\partial^2}{\partial z^2} \nabla^2 T - \nabla^4 T \right) = - \frac{\alpha E}{1-\sigma} \sum_{k=0}^3 [z^k \nabla_1^4 T_k + k(k-1)z^{k-2} \nabla_1^2 T_k],$$

so that (4.1) gives the identity in  $z$ ,

$$\begin{aligned} \sum_{j=0}^n [z^j \nabla_1^4 a_j + 2j(j-1)z^{j-2} \nabla_1^2 a_j + j(j-1)(j-2)(j-3)a_j z^{j-4}] \\ = - \frac{\alpha E}{1-\sigma} \sum_{k=0}^3 [z^k \nabla_1^4 T_k + k(k-1)z^{k-2} \nabla_1^2 T_k]. \end{aligned}$$

Equating the coefficients of like powers of  $z$  on both sides of this equation gives a system of partial differential equations to be satisfied by the  $a_j$ , namely,

\* This is not an assumption since we prescribe  $T$ . Besides the space variables  $x$  and  $y$ ,  $T_k$  may contain the time variable as a parameter.

<sup>†</sup> Nádai, loc. cit., p. 265.

<sup>‡</sup> S. Timoshenko, loc. cit., p. 207.

$$(4.6) \quad \nabla_1^4 \alpha_k + 2\nabla_1^2 \alpha_{k+2} + \alpha_{k+4} = \nabla_1^2 (\nabla_1^2 \tau_k + \tau_{k+2}), \quad k = 0, 1, 2, \dots, n,$$

where

$$\alpha_k \equiv k!a_k, \quad \tau_k \equiv k!cT_k, \quad \alpha_i \equiv 0, \text{ if } i > n, \quad \tau_j \equiv 0, \text{ if } j > 3,$$

and

$$c \equiv -\frac{\alpha E}{1 - \sigma}.$$

Two important observations regarding the structure of the system (4.6) are in order:

(i) The system can be broken up into two independent systems of identical forms, one of which depends on the  $\alpha_k$  with even subscripts, while the other depends on the  $\alpha_k$  with odd subscripts.

(ii) Every  $\alpha_k$  in the first system can be expressed explicitly in terms of  $\alpha_0$  and  $\alpha_2$ , whereas  $\alpha_1$  and  $\alpha_3$  determine every  $\alpha_k$  of the second system. The last assertion becomes clear when the first  $n-3$  of equations (4.6) are solved for  $\alpha_{k+4}$  to give

$$(4.7) \quad \alpha_{k+4} = \nabla_1^4 (\tau_k - \alpha_k) + \nabla_1^2 (\tau_{k+2} - 2\alpha_{k+2}), \quad k = 0, 1, 2, \dots, n-4.$$

Inasmuch as the system of equations associated with the even subscripts on the  $\alpha_k$ 's is of the same form as that for the odd ones, it will suffice to consider only one of them.

Thus, setting  $k=0$  in (4.7) gives

$$\alpha_4 = \nabla_1^4 (\tau_0 - \alpha_0) + \nabla_1^2 (\tau_2 - 2\alpha_2).$$

Substituting this value of  $\alpha_4$  in  $\alpha_6$  obtained from (4.7) by taking  $k=2$  gives

$$\alpha_6 = -2\nabla_1^6 (\tau_0 - \alpha_0) - \nabla_1^4 (\tau_2 - 3\alpha_2).$$

Introducing these expressions for  $\alpha_6$  and  $\alpha_4$  in the right-hand member of (4.7) with  $k=4$  gives  $\alpha_8$  entirely in terms of  $\alpha_0$  and  $\alpha_2$ . The continuation of this process of successive substitutions gives the formula

$$(4.8) \quad \alpha_{2k} = (-1)^k [(k-1)\nabla_1^{2k} (\tau_0 - \alpha_0) + \nabla_1^{2k-2} (\tau_2 - k\alpha_2)],$$

$$k = 2, 3, \dots, N,$$

where  $N \equiv (n/2)$ , the greatest integer contained in  $n/2$ . Thus, it is merely necessary to obtain the solutions for  $\alpha_0$  and  $\alpha_2$  in order to determine completely the remaining  $\alpha_{2k}$ . The corresponding result for the odd  $\alpha_k$  is

$$(4.9) \quad \alpha_{2k+1} = (-1)^k [(k-1)\nabla_1^{2k} (\tau_1 - \alpha_1) + \nabla_1^{2k-2} (\tau_3 - k\alpha_3)],$$

$$k = 2, 3, \dots, M,$$

where  $M \equiv ((n-1)/2)$ , the greatest integer contained in  $(n-1)/2$ .

5. **The determination of  $a_0, a_1, a_2, a_3$ .** The boundary conditions (4.2) and (4.3) impose some restrictions on the functions  $\alpha_k$ , and the nature of these restrictions will be investigated next. Substituting (4.4) in (4.2) and (4.3) gives

$$\begin{aligned} a_0 + a_1 h + a_2 h^2 + a_3 h^3 &= -[a_4 h^4 + \cdots + a_n h^n], \\ a_0 - a_1 h + a_2 h^2 - a_3 h^3 &= -[a_4 h^4 - \cdots + (-1)^n a_n h^n], \\ a_1 + 2a_2 h + 3a_3 h^2 &= -[4a_4 h^3 + \cdots + na_n h^{n-1}], \\ a_1 - 2a_2 h + 3a_3 h^2 &= -[-4a_4 h^3 + \cdots + (-1)^{n-1} na_n h^{n-1}]. \end{aligned}$$

The determinant of the coefficients of the  $a_j$  in the left-hand members of this system of equations is equal to  $-16h^4$ ; so one can solve for  $a_j$ , ( $j=0, 1, 2, 3$ ), in terms of the remaining ones. If  $n \geq 4$ , the solution for  $a_0$  is\*

$$a_0 = \frac{1}{16h^4} \sum_{j=4}^n a_j h^{j-1} \begin{vmatrix} h & h & h^2 & h^3 \\ (-1)^j h & -h & h^2 & -h^3 \\ j & 1 & 2h & 3h^2 \\ (-1)^{j-1} j & 1 & -2h & 3h^2 \end{vmatrix}.$$

It is clear that the value of the coefficient of  $h^{j-1}$  corresponding to odd values of  $j$  is zero. Setting  $j=2k$  gives, after some elementary reductions,

$$(5.1) \quad \alpha_0 = a_0 = \sum_{k=2}^N (k-1) h^{2k} a_{2k} = \sum_{k=2}^N \frac{(k-1)}{(2k)!} h^{2k} \alpha_{2k}.$$

Similarly solving for  $a_2, a_1$ , and  $a_3$  one obtains

$$(5.2) \quad \alpha_2 = 2a_2 = -2 \sum_{k=2}^N k h^{2k-2} a_{2k} = -2 \sum_{k=2}^N \frac{k}{(2k)!} h^{2k-2} \alpha_{2k},$$

$$(5.3) \quad \alpha_1 = a_1 = \sum_{k=2}^M (k-1) h^{2k} a_{2k+1} = \sum_{k=2}^M \frac{k-1}{(2k+1)!} h^{2k} \alpha_{2k+1},$$

$$(5.4) \quad \alpha_3 = 6a_3 = -6 \sum_{k=2}^M k h^{2k-2} a_{2k+1} = -6 \sum_{k=2}^M \frac{k}{(2k+1)!} h^{2k-2} \alpha_{2k+1}.$$

It is thus seen that the equations arising from the boundary conditions also fall into two independent groups associated with even and odd subscripts on  $a_j$ , respectively. The important consequence of this observation lies in the fact that one is led to two independent systems of differential equations and boundary conditions, where the functions  $a_{2k}$  are determined by  $T_0$  and  $T_2$ , and the  $a_{2k+1}$  by  $T_1$  and  $T_3$ .

\* If  $n < 4$ ,  $a_0 = a_1 = a_2 = a_3 = 0$ . See in this connection §7 and §8.

Inasmuch as the two systems of equations are of the same form, the discussion to follow is confined to the system depending on the even subscripts.

Referring to (4.6) and (4.8), and remembering that  $n \geq 4$ , one sees that this latter system is

$$(5.5) \quad \alpha_{2k} = (-1)^k [(k-1)\nabla_1^{2k}(\tau_0 - \alpha_0) + \nabla_1^{2k-2}(\tau_2 - k\alpha_2)],$$

$$(5.6) \quad \nabla_1^2(\nabla_1^2\alpha_{2N-2} + 2\alpha_{2N}) = \nabla_1^4\tau_{2N-2}, \quad k = 2, 3, \dots, N,$$

$$(5.7) \quad \nabla_1^4\alpha_{2N} = 0.$$

The right-hand member in (5.6) vanishes if  $N > 2$ . Consider first the system (5.5), (5.6), (5.7) when  $N \geq 3$ . Using (5.5) with  $k = N$  and  $k = N-1$ , and substituting the resulting values for  $\alpha_{2k}$  in (5.6) and (5.7) give, after some algebraic reductions,

$$(5.8) \quad \nabla_1^{2N}[N\nabla_1^2\alpha_0 + (N+1)\alpha_2] = \nabla_1^{2N}(N\nabla_1^2\tau_0 + \tau_2),$$

$$(5.9) \quad \nabla_1^{2N+2}[(N-1)\nabla_1^2\alpha_0 + N\alpha_2] = \nabla_1^{2N+2}[(N-1)\nabla_1^2\tau_0 + \tau_2].$$

The differential equations for  $\alpha_0$  and  $\alpha_2$  are obtained by operating on (5.8) with  $\nabla_1^2$  and solving for  $\nabla_1^{2N+4}\alpha_0$  and  $\nabla_1^{2N+2}\alpha_2$ . The result is

$$(5.10) \quad \nabla_1^{2N+2}\alpha_2 = \nabla_1^{2N+2}\tau_2,$$

$$(5.11) \quad \nabla_1^{2N+4}\alpha_0 = \nabla_1^{2N+2}(\nabla_1^2\tau_0 - \tau_2).$$

These equations will be solved by a device of constructing a sequence of differential equations of lower orders which  $\alpha_0$  and  $\alpha_2$  must satisfy.

Substitution of (5.5) in the boundary conditions (5.1) and (5.2) gives

$$(5.12) \quad \alpha_0 = \sum_{k=2}^N \frac{(-1)^k(k-1)}{(2k)!} h^{2k} [(k-1)\nabla_1^{2k}(\tau_0 - \alpha_0) + \nabla_1^{2k-2}(\tau_2 - k\alpha_2)],$$

$$(5.13) \quad \alpha_2 = - \sum_{k=2}^N \frac{(-1)^k 2k}{(2k)!} h^{2k-2} [(k-1)\nabla_1^{2k}(\tau_0 - \alpha_0) + \nabla_1^{2k-2}(\tau_2 - k\alpha_2)].$$

Operating on (5.12) with  $\nabla_1^{2N}$  produces

$$\begin{aligned} \nabla_1^{2N}\alpha_0 &= \sum_{k=2}^N \frac{(-1)^k(k-1)}{(2k)!} h^{2k} [(k-1)\nabla_1^{2k+2N}(\tau_0 - \alpha_0) + \nabla_1^{2k+2N-2}(\tau_2 - k\alpha_2)] \\ &= \sum_{k=2}^N \frac{(-1)^k(k-1)}{(2k)!} h^{2k} [(k-1)\nabla_1^{2k+2N}(\tau_0 - \alpha_0) + k\nabla_1^{2k+2N-2}(\tau_2 - \alpha_2) \\ &\quad - (k-1)\nabla_1^{2k+2N-2}\tau_2] \\ &= \sum_{k=2}^N \frac{(-1)^k(k-1)}{(2k)!} h^{2k} \nabla_1^{2k-4} \{ (k-1)[\nabla_1^{2N+2}(\nabla_1^2\tau_0 - \tau_2) - \nabla_1^{2N+4}\alpha_0] \\ &\quad + k[\nabla_1^{2N+2}\tau_2 - \nabla_1^{2N+2}\alpha_2] \}, \end{aligned}$$

which vanishes since the terms in the brackets vanish by (5.10) and (5.11). Thus

$$\nabla_1^{2N}\alpha_0 = 0.$$

In precisely the same way it is shown that

$$\nabla_1^{2N}\alpha_2 = 0.$$

Referring to (5.8) and (5.9) one sees that these last two equations demand that

$$(5.14) \quad N\nabla_1^{2N+2}\tau_0 + \nabla_1^{2N}\tau_2 = 0, \quad (N-1)\nabla_1^{2N+4}\tau_0 + \nabla_1^{2N+2}\tau_2 = 0.$$

Therefore\*

$$(5.15) \quad \nabla_1^{2N+4}\tau_0 = 0, \quad \nabla_1^{2N+2}\tau_2 = 0.$$

Again calculating  $\nabla_1^{2N-2}\alpha_0$  and  $\nabla_1^{2N-2}\alpha_2$  from (5.12) and (5.13), and taking account of the relations just found furnishes two equations for  $\alpha_0$  and  $\alpha_2$  of lower order than the preceding ones, namely,

$$\nabla_1^{2N-2}\alpha_0 = \frac{h^4}{4!} \nabla_1^{2N}(\nabla_1^2\tau_0 + \tau_2), \quad \nabla_1^{2N-2}\alpha_2 = -\frac{h^2}{3!} \nabla_1^{2N}(\nabla_1^2\tau_0 + \tau_2).$$

The result of operating on (5.12) and (5.13) with  $\nabla_1^{2N-4}$  and noting the two equations just found is

$$\begin{aligned} \nabla_1^{2N-4}\alpha_0 &= \frac{h^4}{4!} \nabla_1^{2N-2}(\nabla_1^2\tau_0 + \tau_2) + \frac{h^6}{6!} \nabla_1^{2N}(6\nabla_1^2\tau_0 + 8\tau_2), \\ \nabla_1^{2N-4}\alpha_2 &= -\frac{h^2}{3!} \nabla_1^{2N-2}(\nabla_1^2\tau_0 + \tau_2) - \frac{h^4}{6!} \nabla_1^{2N}(28\nabla_1^2\tau_0 + 34\tau_2). \end{aligned}$$

A little reflection will show that a continuation of this process of operating successively on  $\alpha_0$  and  $\alpha_2$  with  $\nabla_1^{2N-2p}$ , ( $p=0, 1, 2, \dots, N$ ), will yield, after  $N$  operations, expressions for  $\alpha_0$  and  $\alpha_2$  of the type

$$(5.16) \quad \alpha_0 = \sum_{n=2}^{N+1} A_{2n} h^{2n}, \quad \alpha_2 = \sum_{n=2}^{N+1} B_{2n-2} h^{2n-2},$$

where

$$(5.17) \quad \begin{aligned} A_{2n} &= \lambda_{2n} \nabla_1^{2n}\tau_0 + \mu_{2n} \nabla_1^{2n-2}\tau_2, \\ B_{2n} &= \rho_{2n} \nabla_1^{2n+2}\tau_0 + \sigma_{2n} \nabla_1^{2n}\tau_2, \end{aligned}$$

in which  $\lambda, \mu, \rho$ , and  $\sigma$  are constants.

\* For significance of this see §8 below.

The device just outlined can be used successfully to determine the  $\alpha_0$  and  $\alpha_2$  for any  $N \geq 3$ , but it is simpler to calculate the functions  $A_{2n}$  and  $B_{2n}$  by the method of undetermined coefficients since the forms of  $\alpha_0$  and  $\alpha_2$  are now known. This method is followed in the next section where a set of compact formulas is developed for the determination of these functions.

An argument in every respect similar to that just used shows that if  $M \geq 3$ ,

$$\begin{aligned} \alpha_1 &= \sum_{n=2}^{M+1} \bar{A}_{2n} h^{2n}, \\ \alpha_3 &= \sum_{n=2}^{M+1} \bar{B}_{2n-2} h^{2n-2}, \end{aligned} \quad (5.18)$$

where  $\bar{A}_{2n}$  and  $\bar{B}_{2n}$  are of the form (5.17) with  $\tau_0$  and  $\tau_2$  replaced by  $\tau_1$  and  $\tau_3$ , respectively. The functions  $\tau_1$  and  $\tau_3$  satisfy the equations

$$\begin{aligned} \nabla_1^{2M}(M\nabla_1^2\tau_1 + \tau_3) &= 0, \\ \nabla_1^{2M+4}\tau_1 &= 0, \\ \nabla_1^{2M+2}\tau_3 &= 0, \end{aligned} \quad (5.19)$$

which correspond to (5.14) and (5.15).

**6. Recursion formulas for  $A_k$  and  $B_k$ .** Substituting the expressions for  $\alpha_0$  and  $\alpha_2$  from (5.16) in the right-hand member of (5.12) and rearranging gives

$$\begin{aligned} \alpha_0 &= \sum_{k=2}^N \frac{(-1)^k(k-1)}{(2k)!} h^{2k} \left[ (k-1)\nabla_1^{2k} \left( \tau_0 - \sum_{j=2}^{N+1} A_{2j} h^{2j} \right) \right. \\ &\quad \left. + \nabla_1^{2k-2} \left( \tau_2 - k \sum_{j=1}^N B_{2j} h^{2j} \right) \right] \\ &= \sum_{k=2}^N \frac{(-1)^k(k-1)}{(2k)!} \left\{ [(k-1)\nabla_1^{2k-2}(\nabla_1^2\tau_0 - \tau_2) + k\nabla_1^{2k-2}\tau_2] h^{2k} \right. \\ &\quad - k\nabla_1^{2k-2}B_2 h^{2k+2} - \sum_{j=2}^N h^{2k+2j} \nabla_1^{2k-2} [(k-1)\nabla_1^2 A_{2j} + kB_{2j}] \\ &\quad \left. - (k-1)\nabla_1^{2k} A_{2N+2} h^{2N+2k+2} \right\}. \end{aligned}$$

This expression can be written as

$$(6.1) \quad \alpha_0 = \sum_{k=2}^N \frac{(-1)^{k-1}(k-1)}{(2k)!} \sum_{j=0}^{N+1} h^{2k+2j} \nabla_1^{2k-2} [(k-1)\nabla_1^2 A_{2j} + kB_{2j}],$$

where the following definitions are introduced:

$$(6.2) \quad \nabla_1^{2k} A_0 = \nabla_1^{2k-2} (-\nabla_1^2 \tau_0 + \tau_2), \quad \nabla_1^{2k-2} B_0 = -\nabla_1^{2k-2} \tau_2, \\ \nabla_1^{2k} A_2 = 0, \quad \nabla_1^{2k-2} B_{2N+2} = 0.$$

A similar substitution in the right-hand member of (5.13) gives

$$(6.3) \quad \alpha_2 = \sum_{k=2}^N \frac{(-1)^{k-2} k^{N+1}}{(2k)!} \sum_{j=0}^{N+1} h^{2k+2j-2} \nabla_1^{2k-2} [(k-1) \nabla_1^2 A_{2j} + k B_{2j}].$$

In order to collect the coefficients of powers of  $h$ , set  $k+j=r$ ; then (6.1) and (6.3) become\*

$$(6.4) \quad \alpha_0 = \sum_{r=2}^{2N+1} h^{2r} \sum_{j=0}^{r-2} (-1)^{r-j-1} \frac{r-j-1}{(2r-2j)!} \nabla_1^{2r-2j-2} [(r-j-1) \nabla_1^2 A_{2j} + (r-j) B_{2j}], \\ \alpha_2 = \sum_{r=2}^{2N+1} h^{2r-2} \sum_{j=0}^{r-2} (-1)^{r-j} \frac{2(r-j)}{(2r-2j)!} \nabla_1^{2r-2j-2} [(r-j-1) \nabla_1^2 A_{2j} + (r-j) B_{2j}].$$

Equating the coefficients of like powers of  $h$  in (5.16) and (6.4) gives the desired recursion formulas

$$(6.5) \quad A_{2r} = \sum_{j=0}^{r-2} (-1)^{r-j-1} \frac{r-j-1}{(2r-2j)!} \nabla_1^{2r-2j-2} [(r-j-1) \nabla_1^2 A_{2j} + (r-j) B_{2j}], \\ B_{2r-2} = \sum_{j=0}^{r-2} (-1)^{r-j} \frac{2(r-j)}{(2r-2j)!} \nabla_1^{2r-2j-2} [(r-j-1) \nabla_1^2 A_{2j} + (r-j) B_{2j}], \\ r = 2, 3, \dots, N+1.$$

Referring to (6.2) it is clear that the functions  $A_{2k}$  and  $B_{2k}$  can be expressed for any  $N \geq 3$  in terms of the prescribed functions  $\tau_0$  and  $\tau_2$ , so that  $\alpha_0$  and  $\alpha_2$ , and hence  $\alpha_{2k}$ , are completely determined.

The recursion formulas for  $\bar{A}_{2n}$  and  $\bar{B}_{2n}$  in (5.18), deduced by precisely the same method, are

$$(6.6) \quad \bar{A}_{2r} = \sum_{j=0}^{r-2} (-1)^{r-j-1} \frac{r-j-1}{(2r-2j+1)!} \nabla_1^{2r-2j-2} [(r-j-1) \nabla_1^2 \bar{A}_{2j} + (r-j) \bar{B}_{2j}], \\ \bar{B}_{2r-2} = \sum_{j=0}^{r-2} (-1)^{r-j} \frac{6(r-j)}{(2r-2j+1)!} \nabla_1^{2r-2j-2} [(r-j-1) \nabla_1^2 \bar{A}_{2j} + (r-j) \bar{B}_{2j}],$$

where

$$\nabla_1^{2k} \bar{A}_0 = \nabla_1^{2k-2} (-\nabla_1^2 \tau_1 + \tau_3), \quad \nabla_1^{2k-2} \bar{B}_0 = -\nabla_1^{2k-2} \tau_3, \\ \nabla_1^{2k} \bar{A}_2 = 0, \quad \nabla_1^{2k-2} \bar{B}_{2M+2} = 0.$$

\* This amounts to rearranging the finite double sums in (6.1) and (6.2) so that summations proceed first along the diagonals.

7. **Special case**  $N=2$ ,  $M=2$ . The determination of  $Z_s$  was carried out under the assumption that  $N \geq 3$  and  $M \geq 3$ , so that the degree of the polynomial (4.4) was assumed to be greater than 5. It follows from the boundary conditions\* that unless  $n > 3$  the only possible solution for  $Z_s$ , of the form (4.4), is  $Z_s = 0$ . It remains to consider the case when  $n=4$  or  $n=5$ , that is,  $N=M=2$ .

Referring to (5.5), (5.6), and (5.7), one sees that the system of equations associated with even subscripts in this case is

$$(7.1) \quad \alpha_4 = \nabla_1^4 \tau_0 + \nabla_1^2 \tau_2 - \nabla_1^4 \alpha_0 - 2\nabla_1^2 \alpha_2,$$

$$(7.2) \quad \nabla_1^4 \alpha_2 + 2\nabla_1^2 \alpha_4 = \nabla_1^4 \tau_2,$$

$$(7.3) \quad \nabla_1^4 \alpha_4 = 0,$$

and the boundary conditions (5.1) and (5.2) become

$$(7.4) \quad \alpha_0 = \frac{h^4}{4!} \alpha_4, \quad \alpha_2 = -\frac{h^2}{3!} \alpha_4.$$

From (7.4) and (7.3) it follows that

$$(7.5) \quad \nabla_1^4 \alpha_0 = 0, \quad \nabla_1^4 \alpha_2 = 0.$$

Hence (7.2) gives

$$(7.6) \quad \nabla_1^2 \alpha_4 = \frac{1}{2!} \nabla_1^4 \tau_2.$$

Since  $\nabla_1^4 \alpha_2 = 0$ , (7.6) requires that

$$(7.7) \quad \nabla_1^6 \tau_2 = 0.$$

Calculating  $\nabla_1^4 \alpha_4$  from (7.1) and making use of (7.7) gives

$$(7.8) \quad \nabla_1^8 \tau_0 = 0.$$

But from (5.12) and (5.13),

$$(7.9) \quad \alpha_0 = \frac{h^4}{4!} (\nabla_1^4 \tau_0 + \nabla_1^2 \tau_2 - \nabla_1^4 \alpha_0 - 2\nabla_1^2 \alpha_2),$$

and

$$(7.10) \quad \alpha_2 = -\frac{h^2}{3!} (\nabla_1^4 \tau_0 + \nabla_1^2 \tau_2 - \nabla_1^4 \alpha_0 - 2\nabla_1^2 \alpha_2).$$

Calculating  $\nabla_1^2 \alpha_2$  from (7.10) and noting (7.5) gives

---

\* See the footnote in the beginning of §5.

$$\nabla_1^2 \alpha_2 = \frac{h^2}{3!} (\nabla_1^2 \tau_0 + \nabla_1^4 \tau_2),$$

which upon substitution in the right-hand members of (7.9) and (7.10) gives the desired result:

$$\alpha_0 = \frac{h^4}{4!} f_1, \quad \alpha_2 = -\frac{h^2}{3!} f_1, \quad \alpha_4 = f_1,$$

where

$$f_1 \equiv \left(1 + \frac{h^2}{3} \nabla_1^2\right) (\nabla_1^4 \tau_0 + \nabla_1^2 \tau_2).$$

The expression for  $\alpha_4$  just given is obtained from (7.1).

It will be observed that (7.7) and (7.8) are given by (5.15) with  $N=2$ . It is easily checked with the aid of (7.1) and (7.6) that (5.14) is valid even when  $N=2$ .

The formulas for  $\alpha_1$ ,  $\alpha_3$ , and  $\alpha_5$  are found to be

$$\alpha_1 = \frac{h^4}{5!} f_2, \quad \alpha_3 = -\frac{h^2}{10} f_2, \quad \alpha_5 = f_2,$$

where

$$f_2 \equiv \left(1 + \frac{h^2}{5} \nabla_1^2\right) (\nabla_1^4 \tau_1 + \nabla_1^2 \tau_3).$$

It is easily checked that the equations (5.19) are valid when  $M=2$ .

**8. Critique of method.** The determination of  $Z_s$ , outlined above, leaves nothing to be desired from the standpoint of rigor, provided that the assumption that  $Z_s$  is adequately represented by a polynomial in  $z$ , of sufficiently high degree, is satisfied. The assumption that  $Z_s$  is expressible as an infinite series in powers of  $z$  leads to an infinite system of partial differential equations on the  $a_k(x, y)$  of the form (4.6). This in turn will lead to infinite series developments for the  $a_k$ . An analysis of the behavior of the resulting series will prove exceedingly difficult, and is not attempted here. On the other hand, it will be recalled that a considerable portion of the theory of elasticity is built on the assumption that one is dealing with a class of functions which can be approximated arbitrarily closely by polynomials of sufficiently high degree. If it further be noted that a successful theory of moderately thick plates\* has been developed on the hypothesis that  $Z_s$  is a polynomial of degree 3 in  $z$ , the assumption involved in (4.6) should be regarded as exceedingly

\* Cf. A. E. H. Love, loc. cit., pp. 465-487.

mild. However, it must be noted that the choice of  $N$  is connected with the prescribed temperature function  $T$  via equations (5.14), (5.15), and (5.19). The significance of this connection becomes clear upon reflecting that one cannot hope to satisfy the differential equation (4.1), for an entirely arbitrary  $T$ , if he selects the solution for  $Z_z$  in the form (4.4) where  $n$  is small. For example, if an attempt is made to satisfy (4.1) by assuming a solution of the form (4.4) with  $n=4$  or 5, then the polynomial solution will exist only for a class of temperature functions of the form (4.5) in which  $T_k(x, y) = \tau_k/k!c$  satisfy the equations (5.14), (5.15), and (5.19). Obviously, the class of admissible temperature functions is greatly increased with the increase of the degree  $n$  in (4.4). In fact, if the  $T_k(x, y)$  in (4.5) are polynomials in  $x$  and  $y$ , it is always possible to find a pair of numbers  $N$  and  $M$  so great that

$$\begin{aligned}\nabla_1^{2N+2}T_0 &= 0, & \nabla_1^{2N}T_2 &= 0, \\ \nabla_1^{2M+2}T_1 &= 0, & \nabla_1^{2M}T_3 &= 0.\end{aligned}$$

Such a choice of  $N$  and  $M$  will, certainly, satisfy (5.14), (5.15), and (5.19) identically; hence the exact solution of the form (4.4) can be obtained. The case where the  $T_k(x, y)$  are polynomials in  $x$  and  $y$  presents the most interesting distribution from the point of view of applications.

**9. Linear distribution of temperature.** It is of interest to investigate just how far the behavior of an elastic plate subjected to a nonuniform distribution of temperature departs from the state of generalized plane stress assumed by Nádai and Marguerre. It will suffice to consider the case of linear distribution of temperature of the form

$$T = T_0(x, y) + zT_1(x, y).$$

The calculations will be carried out under the assumption that  $\nabla_1^{10}T_0=0$  and  $\nabla_1^8T_1=0$ . Accordingly, the degree of the polynomial in  $z$  for  $Z_z$  in (4.4) is eight. This choice of  $T$  will certainly be satisfactory if  $T_0$  and  $T_1$  are polynomials in  $x$  and  $y$  of degrees 9 and 7, respectively.

Referring to (6.5) and (6.6) one finds easily that

$$\begin{aligned}A_4 &= \frac{1}{4!} \nabla_1^4 T_0, & B_4 &= -\frac{2}{4!} \nabla_1^4 T_0, & A_6 &= \frac{1}{5!} \nabla_1^6 T_0, \\ B_6 &= -\frac{7}{3 \cdot 5!} \nabla_1^6 T_0, & A_8 &= \frac{41}{3 \cdot 8!} \nabla_1^8 T_0, & B_8 &= -\frac{37}{6 \cdot 7!} \nabla_1^8 T_0, \\ \bar{A}_4 &= \frac{1}{5!} \nabla_1^4 T_1, & \bar{B}_2 &= -\frac{2}{5!} \nabla_1^4 T_1, & \bar{A}_6 &= \frac{22}{5 \cdot 7!} \nabla_1^6 T_1, & \bar{B}_4 &= -\frac{54}{5 \cdot 7!} \nabla_1^6 T_1.\end{aligned}$$

Hence from (5.16) and (5.18) it follows that

$$\begin{aligned}
 a_0 &= \frac{\alpha E}{4!(1-\sigma)} \left( -h^4 \nabla_1^4 T_0 - \frac{h^6}{5} \nabla_1^6 T_0 - \frac{41h^8}{5040} \nabla_1^8 T_0 \right), \\
 a_2 &= \frac{\alpha E}{4!(1-\sigma)} \left( 2h^2 \nabla_1^4 T_0 + \frac{7h^4}{15} \nabla_1^6 T_0 + \frac{37h^6}{1260} \nabla_1^8 T_0 \right), \\
 a_1 &= -\frac{\alpha E}{5!(1-\sigma)} \left( h^4 \nabla_1^4 T_1 + \frac{11h^6}{105} \nabla_1^6 T_1 \right), \\
 a_3 &= \frac{\alpha E}{5!(1-\sigma)} \left( 2h^2 \nabla_1^4 T_1 + \frac{9h^4}{35} \nabla_1^6 T_1 \right).
 \end{aligned}$$

The remaining functions  $a_i$  are readily calculated from (4.7); the results are

$$\begin{aligned}
 a_4 &= -\frac{\alpha E}{4!(1-\sigma)} \left( \nabla_1^4 T_0 + \frac{h^2}{3} \nabla_1^6 T_0 + \frac{13h^4}{360} \nabla_1^8 T_0 \right), \\
 a_6 &= \frac{\alpha E}{4!(1-\sigma)} \left( \frac{1}{15} \nabla_1^6 T_0 + \frac{h^2}{60} \nabla_1^8 T_0 \right), \\
 a_8 &= -\frac{\alpha E}{4!(1-\sigma)} \frac{1}{560} \nabla_1^8 T_0, \\
 a_5 &= -\frac{\alpha E}{5!(1-\sigma)} \left( \nabla_1^4 T_1 + \frac{h^2}{5} \nabla_1^6 T_1 \right), \\
 a_7 &= \frac{\alpha E}{5!(1-\sigma)} \frac{1}{21} \nabla_1^6 T_1.
 \end{aligned}$$

Substituting these coefficients in (4.4) gives

$$\begin{aligned}
 (9.1) \quad Z_z &= \frac{\alpha E}{4!(1-\sigma)} (z^2 - h^2)^2 \left( \nabla_1^4 T_0 - \frac{z}{5} \nabla_1^4 T_1 + \frac{z^2 - 3h^2}{15} \nabla_1^6 T_0 \right. \\
 &\quad \left. + \frac{5z^3 - 11h^2z}{525} \nabla_1^6 T_1 - \frac{9z^4 - 66z^2h^2 + 41h^4}{7!} \nabla_1^8 T_0 \right).
 \end{aligned}$$

If  $n$  is assumed to be greater than 8, the additional terms appearing in the expression for  $Z_z$  do not affect the coefficients of the powers of  $\nabla_1$  just found.

It is clear that a similar analysis, pertaining to unheated plates, subjected to an arbitrary load, can be carried out. In this way one will arrive, from an altogether different point of view, at the solution of the problem discussed by Southwell.\*

**10. Shearing stresses  $Z_x$  and  $Z_y$ .** The determination of the remaining stress components acting in the direction of the thickness of the plate will be

\* R. V. Southwell, loc. cit.

given in this section, the point of interest being that it is possible to obtain these stresses directly from the fundamental equations and without imposing restrictions as severe as those implied by (2.8) and (2.9).

The equation (2.4) with  $i=j=3$  gives

$$(10.1) \quad -\frac{\partial^2 \Theta}{\partial z^2} = (1 + \sigma) \nabla^2 Z_z + \frac{1 + \sigma}{1 - \sigma} \alpha E \nabla^2 T + \alpha E \frac{\partial^2 T}{\partial z^2},$$

and from (2.5) one obtains

$$(10.2) \quad -\frac{\partial^2 \Theta}{\partial z^2} = \nabla^2 \Theta + \frac{2\alpha E}{1 - \sigma} \nabla^2 T.$$

Equating (10.1) and (10.2) and simplifying gives

$$(10.3) \quad \nabla^2 \Theta = (1 + \sigma) \nabla^2 Z_z - \alpha E \nabla^2 T.$$

Setting  $i=2$  and  $j=3$  in equations (2.4) yields

$$(10.4) \quad \nabla^2 Z_y = -\frac{1}{1 + \sigma} \frac{\partial^2 \Theta}{\partial y \partial z} - \frac{\alpha E}{1 + \sigma} \frac{\partial^2 T}{\partial y \partial z}.$$

Taking  $\nabla^2$  of both sides of (10.4) and substituting from (10.3) furnishes the relation

$$(10.5) \quad \nabla^2 \nabla^2 Z_y = -\nabla^2 \frac{\partial^2 Z_z}{\partial y \partial z}.$$

Hence

$$(10.6) \quad \nabla^2 Z_y = -\frac{\partial^2 Z_z}{\partial y \partial z} + \phi(x, y, z),$$

where  $\phi$  is a harmonic function.

But the last one of the conditions (4.2) demands

$$Z_y(x, y, \pm h) = 0,$$

so that

$$\nabla^2 Z_y = 0 \quad \text{on } z = \pm h.$$

Thus, it follows from (10.6) that  $\phi(x, y, z)$  satisfies the conditions

$$\phi(x, y, \pm h) = \left( \frac{\partial^2 Z_z}{\partial y \partial z} \right)_{z=\pm h}.$$

However, the function  $Z_z$  has been determined, and it follows from the foregoing that its form is

$$Z_z = (z^2 - h^2)^2 \sum_{j=0}^k a_j(x, y) z^j,$$

so that

$$\left( \frac{\partial^2 Z_z}{\partial y \partial z} \right)_{z=\pm h} = 0.$$

Thus the harmonic function  $\phi(x, y, z)$  vanishes on the faces of the plate, and it follows that in a small region about the middle of the plate it cannot differ very much from zero. Hence, as a first approximation, one can choose  $\phi = 0$ . By assuming the solution for  $\phi$  in the form

$$\phi = \sum_{j=0}^n \phi_j(x, y) z^j$$

and following the method of §4, one can improve upon the first approximation,\* but it will be assumed, for the present, that  $\phi = 0$  gives a satisfactory expression for  $\nabla^2 Z_y$  in (10.6).

Therefore the conditions on  $Z_y$  become as follows:

$$(10.7) \quad \nabla^2 Z_y = - \frac{\partial^2 Z_z}{\partial y \partial z},$$

where

$$Z_y = 0 \quad \text{on} \quad z = \pm h.$$

Observing that  $\nabla^2 Z_y = \nabla^2 Z_y + \partial^2 Z_y / \partial z^2$ , and making use of (10.7), one is enabled to write (10.4) in the form

$$(10.8) \quad \frac{\partial^2 Z_y}{\partial z^2} = - \frac{1}{1 + \sigma} \frac{\partial^2 \Theta}{\partial y \partial z} - \frac{\alpha E}{1 + \sigma} \frac{\partial^2 T}{\partial y \partial z} + \frac{\partial^2 Z_z}{\partial y \partial z}.$$

On the other hand, the differentiation of (3.2) with respect to  $y$  gives

$$(10.9) \quad \frac{\partial}{\partial y} \nabla^2 w = - 2\alpha \frac{\partial^2 T}{\partial y \partial z} + \frac{\sigma - 1}{E} \frac{\partial^2 \Theta}{\partial y \partial z} - \frac{1 + \sigma}{E} \frac{\partial^2 Z_z}{\partial y \partial z}.$$

Eliminating  $\partial^2 \Theta / \partial y \partial z$  between (10.8) and (10.9) produces

$$(10.10) \quad \frac{\partial^2 Z_y}{\partial z^2} = \frac{E}{1 - \sigma^2} \frac{\partial}{\partial y} \nabla^2 w + \frac{\alpha E}{1 - \sigma} \frac{\partial^2 T}{\partial y \partial z} + \frac{2 - \sigma}{1 - \sigma} \frac{\partial^2 Z_z}{\partial y \partial z}.$$

If it be assumed that  $w(x, y, z)$  can be replaced by  $w(x, y, 0)$ , then the

\* The process here will be much less involved since  $\phi$  satisfies Laplace's equation.

right-hand member of (10.10) becomes a known function of  $x$ ,  $y$ , and  $z$ , since the deflection of the middle plane,  $z=0$ , can be calculated from (3.7). It is to be noted that the assumption that the deflection of the middle surface be nearly the same as that of any plane parallel to the middle surface is not as severe as (2.8).

Thus, replacing  $w$  by  $w_0$  in (10.10), and integrating twice with respect to  $z$  gives

$$Z_y = \frac{Ez^2}{2(1-\sigma^2)} \frac{\partial}{\partial y} \nabla^2 w_0 + \frac{\alpha E}{1-\sigma} \int_0^z \frac{\partial T}{\partial y} dz + \frac{2-\sigma}{1-\sigma} \int_0^z \frac{\partial Z_z}{\partial y} dz + zf_1 + f_2,$$

where  $f_1$  and  $f_2$  are arbitrary functions of  $x$  and  $y$ .

Since  $Z_y$  vanishes on  $z = \pm h$ , the functions  $f_1$  and  $f_2$  are uniquely determined. An elementary calculation shows that

$$\begin{aligned} Z_y = & \frac{E(z^2 - h^2)}{2(1-\sigma^2)} \frac{\partial}{\partial y} \nabla^2 w_0 + c_1 \int_0^z \frac{\partial T}{\partial y} dz + c_2 \int_0^z \frac{\partial Z_z}{\partial y} dz \\ & - \frac{z}{2h} \left( c_1 \int_{-h}^h \frac{\partial T}{\partial y} dz + c_2 \int_{-h}^h \frac{\partial Z_z}{\partial y} dz \right) \\ (10.11) \quad & - \frac{c_1}{2} \left( \int_0^h \frac{\partial T}{\partial y} dz + \int_0^{-h} \frac{\partial T}{\partial y} dz \right) \\ & - \frac{c_2}{2} \left( \int_0^h \frac{\partial Z_z}{\partial y} dz + \int_0^{-h} \frac{\partial Z_z}{\partial y} dz \right), \end{aligned}$$

where

$$c_1 = \frac{\alpha E}{1-\sigma}, \quad c_2 = \frac{2-\sigma}{1-\sigma}.$$

The expression for  $Z_x$  can be obtained in precisely the same way. The resulting formula is (10.11) with  $y$  replaced by  $x$ .

The first term in the right-hand member of (10.11) is recognized as the expression for shear given by the ordinary thin plate theory under the hypotheses (2.8) and (2.9).

**11. Conclusion.** Some interesting deductions regarding the behavior of a heated elastic plate can be drawn immediately from (3.7). The right-hand member of (3.7) can be regarded as a fictitious normal load  $p(x, y)$  which produces the same deflection in an unheated plate as that caused by the thermal stresses. Now, if the  $T_k$  are constants so that the temperature is a function of  $z$  alone, the right-hand member of (3.7) vanishes and  $w_0$  satisfies

$$\nabla_1^4 w_0 = 0.$$

If the plate is clamped,

$$w_0 = \frac{\partial w_0}{\partial n} = 0$$

along the rigid support, and the only possible solution of the system is  $w_0 = 0$ . Hence a clamped plate under the action of thermal stresses alone will remain plane. The same conclusion is reached upon assuming a linear distribution of temperature of the form  $T = T_0(x, y) + zT_1(x, y)$ , where  $T$  is steady. However, a simply supported plate will buckle under the action of thermal stresses. An investigation pertaining to the action of such plates, based on the theory developed in this paper, will be published elsewhere.

UNIVERSITY OF WISCONSIN,  
MADISON, WIS.

## GEOMETRY OF A SURFACE IN THE NEIGHBORHOOD OF A SPINE\*

BY

V. G. GROVE

1. **Introduction.** In this paper we propose to continue the study begun by Bompiani,† Lane, and others of curves and surfaces in the neighborhood of a singularity. The curves and surfaces previously studied possessed a tangent line or plane at the singularity. We shall discuss from a projective point of view the geometry of a surface in the neighborhood of a singularity we have called a spine by making use of certain osculants of a plane curve at a cusp.

We shall say that a point  $O$  on a surface  $S$  is a *spine* if it is a conical point such that the cone at the conical point degenerates into two planes. A spine may also be defined as a point  $O$  through which there exists a line  $t$  such that every nonspecialized plane section of  $S$  through  $t$  has an ordinary cusp at  $O$  with  $t$  as the cuspidal tangent.

Two cases arise according as the degenerate cone consists of distinct planes or coincident planes. In the first case, which we will call the nonparabolic case, the line  $t$  is unique and is the intersection of the distinct planes. We shall call the line  $t$  the *spinal tangent*. In the second case, which we shall call the parabolic case, the line  $t$  is any line in the pair of coincident planes. We shall call  $t$  a *spinal tangent*, and shall call such a spine  $O$  a *spine of the parabolic type*.

In §§3 and 4 we derive the equations of certain osculating curves of a curve at a cusp. In particular, we define order of contact of an algebraic curve with a given curve at its cusp.

In §2 we derive a simple form of the equation of a surface  $S$  with a spine at  $O$  in the nonparabolic case, and then use the osculants for a curve at a cusp to study the surface

In the last section we briefly discuss the surface in the neighborhood of a spine of the parabolic type.

2. **The canonical form of the equation of  $S$ .** Let the surface  $S$  have a spine of the nonparabolic type at  $O$ . Let the nonhomogeneous projective co-

\* Presented to the Society, April 10, 1937; received by the editors January 24, 1938.

† E. Bompiani, *Per lo studio proiettivo-differenziale delle singolarità*, Bollettino dell'Unione Matematica, vol. 5 (1926), p. 118; E. P. Lane, *The neighborhood of a sextactic point on a plane curve*, Duke Mathematical Journal, vol. 1 (1935), p. 287; T. L. Downs, *Asymptotic and principal directions at a planar point of a surface*, Duke Mathematical Journal, vol. 1 (1935), p. 316; V. G. Grove, *Differential geometry of a surface at a planar point*, these Transactions, vol. 40 (1936), p. 155.

ordinates of a point  $P$  be  $(x, y, z)$ . If the planes  $y=0, z=0$  be chosen as the distinct planes of which the degenerate cone consists, and the point  $(0, 0, 0)$  be chosen at the spine, we may write the equation of the surface in the form

$$(1) \quad yz = A_3 + B_2x + C_1x^2 + D_0x^3 + A_4 + B_3x + C_2x^2 + D_1x^3 + E_0x^4 + \dots,$$

wherein  $A_i, B_i, C_i, \dots$  are homogeneous functions of degree  $i$  in  $y$  and  $z$  only.

By transformations of the form

$$x = x' + \alpha y' + \beta z', \quad y = y', \quad z = z';$$

$$x' = \frac{\bar{x}}{1 + \lambda \bar{x} + \mu \bar{y} + \nu \bar{z}}, \quad y' = \frac{\bar{y}}{1 + \lambda \bar{x} + \mu \bar{y} + \nu \bar{z}}, \quad z' = \frac{\bar{z}}{1 + \lambda \bar{x} + \mu \bar{y} + \nu \bar{z}};$$

$$\bar{x} = q\xi, \quad \bar{y} = r\eta, \quad \bar{z} = s\zeta,$$

we may reduce equation (1) to the following canonical form

$$(2) \quad \eta\zeta = A_3 + B_2\xi + C_1\xi^2 + D_0\xi^3 + A_4 + B_3\xi + C_2\xi^2 + D_1\xi^3 + E_0\xi^4 + \dots,$$

wherein

$$\begin{aligned} A_3 &= A_{30}\eta^3 + A_{21}\eta^2\zeta + A_{12}\eta\zeta^2 + A_{03}\zeta^3, \\ B_2 &= -A_{30}\eta^2 + B_{11}\eta\zeta - A_{03}\zeta^2, \\ C_1 &= 0, \\ D_0 &= 1, \\ A_4 &= A_{40}\eta^4 + A_{31}\eta^3\zeta + A_{22}\eta^2\zeta^2 + A_{13}\eta\zeta^3 + A_{04}\zeta^4, \\ B_3 &= B_{30}\eta^3 + B_{21}\eta^2\zeta + B_{12}\eta\zeta^2 + B_{03}\zeta^3, \\ C_2 &= C_{20}\eta^2 + C_{11}\eta\zeta + C_{02}\zeta^2, \\ D_1 &= -A_{21}\eta - A_{12}\zeta, \\ E_0 &= -B_{11}, \\ &\dots \end{aligned} \tag{3}$$

We shall call the tetrahedron of reference giving rise to the equation (2) *the canonical tetrahedron*. Every coefficient appearing in (2) is an absolute invariant of  $S$  at  $O$ .

We find it convenient to introduce homogeneous projective coordinates by the formulas

$$\xi = \frac{x_1}{x_4}, \quad \eta = \frac{x_2}{x_4}, \quad \zeta = \frac{x_3}{x_4}.$$

**3. Canonical form of the equation of a curve with a cusp.** In this section we shall derive a canonical form of the equation of a curve with a cusp, and

define certain osculating curves of the curve. We find these osculants useful in discussing a surface in the neighborhood of a spine.

Let  $(\xi, \eta)$  be the nonhomogeneous projective coordinates of a point  $P$  in a plane. Let the point  $(0, 0)$  be chosen at the cusp  $O$  of the curve  $C$ , and let the side  $\eta=0$  of the triangle of reference be the cuspidal tangent of  $C$  at  $O$ . By proper choice of the side  $\xi=0$ , the equation of the curve  $C$  may be written in the form

$$(4) \quad \eta^2 = \alpha_{30}\xi^3 + \alpha_{12}\xi\eta^2 + \alpha_{03}\eta^3 + \alpha_{40}\xi^4 + \alpha_{31}\xi^3\eta + \alpha_{22}\xi^2\eta^2 + \alpha_{13}\xi\eta^3 + \alpha_{04}\eta^4 \\ + \alpha_{50}\xi^5 + \alpha_{41}\xi^4\eta + \alpha_{32}\xi^3\eta^2 + \alpha_{23}\xi^2\eta^3 + \alpha_{14}\xi\eta^4 + \alpha_{05}\eta^5 + \dots, \alpha_{30} \neq 0.$$

If on the equation (4) of  $C$  we make, in order, the transformations

$$(5) \quad \xi = \frac{\bar{x}}{1 + p\bar{x} + q\bar{y}}, \quad \eta = \frac{\bar{y}}{1 + p\bar{x} + q\bar{y}}, \\ \bar{x} = rx, \quad \bar{y} = sy,$$

wherein  $p, q, r, s$  are defined by the expressions

$$(6) \quad p\alpha_{30} = \alpha_{40} + \alpha_{31}\alpha_{12}, \quad q\alpha_{30} = \alpha_{31} + \alpha_{30}\alpha_{03}, \\ \alpha_{30}r^3 = s^2, \quad r(\alpha_{22}\alpha_{30} + \alpha_{50}) = s(\alpha_{13}\alpha_{30} + \alpha_{41}),$$

we reduce (4) to the following canonical form

$$(7) \quad y^2 = x^3 + a_{12}xy^2 + a_{03}y^3 - a_{12}x^4 - a_{03}x^3y + a_{22}x^2y^2 + a_{13}xy^3 + a_{04}y^4 \\ + a_{50}x^5 + a_{41}x^4y + a_{32}x^3y^2 + a_{23}x^2y^3 + a_{14}xy^4 + a_{05}y^5 + \dots,$$

wherein  $a_{22} + a_{50} = a_{13} + a_{41} \neq 0$ . We shall speak of the triangle giving rise to the canonical form (7) as *the canonical triangle*.

From the canonical form (7) we derive the equations of  $C$  in the simple parametric form

$$(8) \quad x = t^2, \quad y = t^3 + \alpha_7 t^7 + \alpha_8 t^8 + \alpha_9 t^9 + \alpha_{10} t^{10} + \dots,$$

wherein

$$\alpha_7 = \alpha_8 = \frac{1}{2}(a_{22} + a_{50}), \quad \alpha_9 = \frac{1}{2}[a_{12}(a_{22} + a_{50}) + a_{04} + a_{32} + a_{60}], \dots$$

The transformation of coordinates between the triangle of reference for the equation of the curve  $C$  in the form (4) to the canonical triangle may be written in the form

$$(9) \quad \xi = \frac{rx}{1 + prx + qsy}, \quad \eta = \frac{sy}{1 + prx + qsy},$$

wherein  $p, q, r, s$  are defined by the formulas (6). The inverse of transformation (9) is

$$(10) \quad x = \frac{\xi}{r(1 - p\xi - q\eta)}, \quad y = \frac{\eta}{s(1 - p\xi - q\eta)}.$$

We shall find it convenient to introduce homogeneous projective coordinates by the formulas

$$x = \frac{x_1}{x_3}, \quad y = \frac{x_2}{x_3}.$$

4. **Osculating curves of a curve at a cusp.** Let  $A_{p,q}^n$  be an algebraic curve of order  $n$ , with a multiple point of order  $p+2$  at the cusp  $O$  of  $C$ , two of whose multiple point tangents coincide with the cuspidal tangent  $y=0$  of  $C$  at  $O$ , and the remaining  $p$  of whose tangents do not so coincide with  $y=0$ . Let the number of points of intersection of  $A_{p,q}^n$  with  $C$  coinciding at  $O$  be denoted by  $N$ . We shall say that  $A_{p,q}^n$  has  $q$ -point contact or  $(q-1)$ -order contact with  $C$  at  $O$  if

$$q = N - 2p - 4.$$

In the language of Bompiani we shall say that the curve  $A_{p,q}^n$  represents the neighborhood of the  $(q-1)$ st order of  $C$  at  $O$ , and shall call  $A_{p,q}^n$  an osculating curve of  $C$  at  $O$ .

Writing the most general equation of a cubic with a cusp at  $O$ , and demanding that the power series resulting by the substitution of (8) into that equation shall lack terms in  $t^k$ , ( $k=0, 1, 2, 3, 4, 5$ ), we find the equation of the osculating cubic  $A_{0,6}^3$  to be

$$(11) \quad y^2 = x^3.$$

The equation of the osculating cubic gives a simple geometrical characterization of the canonical triangle. *The vertex  $O_2(0, 1, 0)$  is the inflexion point of the cubic; the side  $x_3=0$  is the inflexion tangent.*

The neighborhoods of  $C$  at  $O$  of the sixth and seventh orders may be represented by the respective curves

$$\begin{aligned} A_{0,7}^4: \quad y^2 &= x^3 + (a_{22} + a_{50})x^2y^2; \\ A_{1,6}^4: \quad y^2(x-y) &= x^4 - x^3y + (a_{22} + a_{50})y^4. \end{aligned}$$

The unit point of the canonical triangle of reference is the intersection of the triple point tangent  $y=x$  of  $A_{1,6}^4$  with the osculating cubic  $A_{0,6}^3$ .

5. **Sections of the surface  $S$  through the spinal tangent.** We now shall use the results of the previous sections in a discussion of a surface in the neighborhood of a spine of the nonparabolic type.

The plane which is given by

$$(12) \quad \zeta = m\eta, \quad m \neq 0,$$

intersects the surface  $S$  whose equation is (2) in a curve with a cusp at  $O$   $(0, 0, 0, 1)$ . The line  $\eta = \zeta = 0$  is the cuspidal tangent for all of these sections. *The planes  $\eta = 0, \zeta = 0$  are the only planes through the spinal tangent which intersect the surface in curves with a triple point at  $O$ .*

The equations of the curve of intersection of the plane (12) with the surface  $S$  are

$$(13) \quad \begin{aligned} \zeta &= m\eta, \\ \eta^2 &= \alpha_{30}\xi^3 + \alpha_{12}\xi\eta^2 + \alpha_{03}\eta^3 + \alpha_{40}\xi^4 + \alpha_{31}\xi^3\eta + \alpha_{22}\xi^2\eta^2 + \alpha_{13}\xi\eta^3 + \alpha_{04}\eta^4 \\ &\quad + \alpha_{50}\xi^5 + \alpha_{41}\xi^4\eta + \dots, \end{aligned}$$

wherein  $\alpha_{30}, \alpha_{12}, \alpha_{03}$  are defined by the formulas

$$(14) \quad \begin{aligned} m\alpha_{30} &= 1, & m\alpha_{12} &= -A_{30} + B_{11}m - A_{03}m^2, \\ m\alpha_{03} &= A_{30} + A_{21}m + A_{12}m^2 + A_{03}m^3, & m\alpha_{40} &= -B_{11}, \\ m\alpha_{31} &= -(A_{21} + A_{12}m), & m\alpha_{22} &= C_{20} + C_{11}m + C_{02}m^2, \\ m\alpha_{13} &= B_{30} + B_{21}m + B_{12}m^2 + B_{03}m^3, \\ m\alpha_{04} &= A_{40} + A_{31}m + A_{22}m^2 + A_{13}m^3 + A_{04}m^4, \\ m\alpha_{50} &= F_{00}, & m\alpha_{41} &= E_{10} + E_{01}m, \dots \end{aligned}$$

The second of equations (13) is of the same form as (4). Hence we may derive the equations of the loci associated with the various neighborhoods of the sections of  $S$  through the spinal tangent by the use of the transformations (9) and (10).

By this method we find that *the equations of the osculating cubic of the section  $C$  by the plane (12) are*

$$(15) \quad \zeta = m\eta, \quad m\eta^2 = \xi^3 - (A_{30} + A_{03}m^2)\xi\eta^2 + (A_{30} + A_{03}m^3)\eta^3.$$

*The locus of the osculating cubic (15) is a cubic surface whose equation is*

$$(16) \quad \eta\zeta = \xi^3 - (A_{30}\eta^2 + A_{03}\zeta^2)\xi + (A_{30}\eta^3 + A_{03}\zeta^3).$$

We shall call this cubic surface *the osculating cubic surface* of  $S$  at  $O$ . The osculating cubic surface of course has a spine at  $O$ .

We note from the form of equations (15) that *the points of inflexion of all of the osculating cubics of the sections of  $S$  through the spinal tangent lie in a plane*. This plane is the face  $\xi = 0$  of the canonical tetrahedron. *The locus of the inflexion points of the osculating cubics is therefore a plane cubic whose equations are*

$$(17) \quad \xi = 0, \quad \eta\zeta = A_{30}\eta^3 + A_{03}\zeta^3.$$

Equations (17) furnish another interpretation of the edges  $\xi = \eta = 0$ ,  $\xi = \zeta = 0$  of the canonical tetrahedron. *The inflexion points of the cubic (17) lie on the edge  $x_1 = x_4 = 0$  of the canonical tetrahedron.*

The inflexion tangent of the osculating cubic (15) intersects the spinal tangent in the point

$$(18) \quad (m, 0, 0, -A_{30} - A_{03}m^2).$$

There is established therefore a (1, 2) correspondence between the points on the spinal tangent and the planes through the spinal tangent. *The vertex  $O_1$  (1, 0, 0, 0) of the canonical tetrahedron is the only point to which correspond planes separating the planes  $\eta = 0$ ,  $\zeta = 0$  harmonically.* There exist on the spinal tangent two points  $D, D_1$  to each of which correspond identical planes. *The vertices  $O$  and  $O_1$  separate  $D, D_1$  harmonically.*

The equations of the inflexional tangent of the osculating cubic (15) are

$$(19) \quad \begin{aligned} \zeta &= m\eta, \\ (A_{30} + A_{03}m^2)\xi + (A_{30} + A_{03}m^2)\eta + m &= 0. \end{aligned}$$

*The locus of this inflexional tangent is the cubic ruled surface*

$$(20) \quad x_2x_3x_4 = -(A_{30}x_2^2 + A_{03}x_3^2)x_1 + A_{30}x_2^3 + A_{03}x_3^3.$$

This ruled surface intersects the edge  $x_3 = x_4 = 0$  of the canonical tetrahedron in the point  $O_1$  and in the point (1, 1, 0, 0). It intersects the edge  $x_2 = x_4 = 0$  in the point  $O_1$  and in the point (1, 0, 1, 0). The point (1, 1, 1, 0) is therefore characterized. *The line joining the spine  $O$  to (1, 1, 1, 0) intersects the osculating cubic surface in the unit point of the canonical tetrahedron.*

We may readily verify that the planes  $\zeta = m\eta$  wherein  $A_{30} - A_{03}m^2 = 0$  intersect the cubic surface (20) in a pair of torsal generators of that surface. *The torsal generators intersect the spinal tangent in the two points  $D$  and  $D_1$  previously described.*

The line  $x_2 = x_3 = 0$  evidently lies on the surface (20). The line whose parametric equations are

$$x_1 = 1 + m, \quad x_2 = 1, \quad x_3 = m, \quad x_4 = -(A_{30} + A_{03}m),$$

also lies on the surface. Hence the surface (20) possesses two straight line directrices (flecnode curves).

*The locus of the triple point tangent of the curve  $A_{1,8}^4$  having eight-point contact with the section of  $S$  at  $O$  is a cubic cone whose equation is*

$$(21) \quad \begin{aligned} B_{30}\eta^3 + (B_{21} + E_{10})\eta^2\zeta + (B_{12} + E_{01})\eta\zeta^3 + B_{03}\zeta^3 \\ = [C_{20}\eta^2 + (C_{11} + F_{00})\eta\zeta + C_{02}\zeta^2]\xi. \end{aligned}$$

The vertex of this cone is of course the spine  $O$ . The spinal tangent is a singular line; the planes

$$C_{20}\eta^2 + (C_{11} + F_{00})\eta\zeta + C_{02}\zeta^2 = 0$$

are the singular tangent planes along the line. *The surface (21) intersects the face  $x_4=0$  of the canonical tetrahedron in a cubic curve which has a double point at  $O_1$ , and whose inflexions lie on the line  $x_1=x_4=0$ .*

Loci associated with other algebraic curves representing neighborhoods of higher order could be considered, but we shall carry this discussion no further.

**6. The parabolic case.** In this section we discuss briefly the geometry of a surface in the neighborhood of a spine of the parabolic type.

By proper choice of the tetrahedron of reference the equation of a surface possessing such a spine may be written in the form

$$(22) \quad \eta^2 = A_3 + B_2\zeta + C_1\zeta^2 + D_0\zeta^3 + A_4 + B_3\zeta + C_2\zeta^2 + D_1\zeta^3 + E_0\zeta^4 + \dots,$$

wherein

$$\begin{aligned} A_3 &= A_{30}\eta^3 + A_{21}\eta^2\zeta - A_{03}\eta\zeta^2 + A_{03}\zeta^3, & A_{03} &\neq 0, \\ B_2 &= B_{20}\eta^2 - A_{03}\zeta^2, \\ C_1 &= 0, \\ D_0 &= 1, \\ (23) \quad A_4 &= A_{40}\eta^4 + A_{31}\eta^3\zeta + A_{22}\eta^2\zeta^2 + A_{13}\eta\zeta^3 + A_{04}\zeta^4, \\ B_3 &= B_{30}\eta^3 + B_{21}\eta^2\zeta + B_{12}\eta\zeta^2 + B_{03}\zeta^3, \\ C_2 &= C_{20}\eta^2 + C_{11}\eta\zeta + C_{02}\zeta^2, \\ D_1 &= -A_{30}\eta - A_{21}\zeta, \\ E_0 &= -B_{20}, \\ &\dots \end{aligned}$$

The plane  $\eta=0$  intersects the surface  $S$  in a curve with a triple point at  $O$ . The triple point tangents are determined by the cubic form

$$\xi^3 - A_{03}\zeta^2(\xi - \zeta) = 0.$$

The interpretation of the condition  $A_{03} \neq 0$  is evident. The line  $\eta=\zeta=0$  is an arbitrary line in the plane  $\eta=0$ .

The section of the surface  $S$  through  $O$   $(0, 0, 0)$  by the plane  $\zeta=m\eta$  has the equations

$$\begin{aligned} \zeta &= m\eta, \\ (24) \quad \eta^2 &= \xi^3 + \alpha_{12}\xi\eta^2 + \alpha_{03}\eta^3 + \alpha_{40}\xi^4 + \alpha_{31}\xi^3\eta + \alpha_{22}\xi^2\eta^2 + \alpha_{13}\xi\eta^3 \\ &\quad + \alpha_{04}\eta^4 + \dots, \end{aligned}$$

wherein

$$\begin{aligned}\alpha_{12} &= B_{20} - A_{03}m^2, \\ \alpha_{03} &= A_{30} + A_{21}m - A_{03}m^2 + A_{03}m^3, & \alpha_{40} &= -B_{20}, \\ \alpha_{31} &= -A_{30} - A_{21}m, & \alpha_{22} &= C_{20} + C_{11}m + C_{02}m^2, \\ \alpha_{13} &= B_{30} + B_{21}m + B_{12}m^2 + B_{03}m^3, \\ \alpha_{04} &= B_{40} + B_{31}m + B_{22}m^2 + B_{13}m^3 + B_{04}m^4, \\ &\dots\end{aligned}$$

If use be made of formula (10) and the definitions of  $p, q, r, s$  occurring therein, we may show that the equations of the osculating cubic of the section (24) of  $S$  are

$$(25) \quad \begin{aligned}\zeta &= m\eta, \\ \eta^2 &= \xi^3 - A_{03}m^2\eta^2[\xi - (1 + m)\eta].\end{aligned}$$

Equations (25) show that the inflexion points of the osculating cubics of all sections through a given arbitrary spinal tangent  $\eta = \zeta = 0$  lie on the fixed plane  $\xi = 0$ .

The locus of the osculating cubic is the cubic surface with the equation

$$(26) \quad \eta^2 = \xi^3 - A_{03}\zeta^2(\xi - \eta - \zeta).$$

The inflexional tangent of the curve (25) intersects the line  $\eta = \zeta = 0$  in a point whose homogeneous coordinates are  $(1, 0, 0, -A_{03}m^2)$ . As in the non-parabolic case, there is thus established a (1, 2) correspondence between the points on  $\eta = \zeta = 0$  and the planes through that spinal tangent. *The point  $O_1$   $(1, 0, 0, 0)$  is the only point on the line to which correspond identical planes. Moreover the planes corresponding to the points on  $\eta = \zeta = 0$  are paired in involution; the double planes of the involution are the planes  $\eta = 0, \zeta = 0$ .*

The line  $\xi = \zeta = 0$  intersects the surface (26) in  $O$  and in the vertex  $O_2$   $(0, 1, 0, 0)$  of the tetrahedron of reference.

The inflexion point of the cubic (25) has the homogeneous coordinates

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = m, \quad x_4 = A_{03}m^2(1 + m).$$

The locus of this point passes through the vertex  $(0, 1, 0, 0)$ , and the tangent to the locus at this vertex is the line  $\xi = \zeta = 0$ .

The inflexional tangent of the osculating cubic (25) has the equations

$$(27) \quad \begin{aligned}\zeta &= m\eta, \\ A_{03}m^2[\xi - (1 + m)\eta] + 1 &= 0.\end{aligned}$$

The locus of this inflexional tangent is the cubic ruled surface

$$x_2^2 x_4 = -A_{03}[x_1 + x_2 - x_3]x_3^2.$$

The edge  $x_3 = x_4 = 0$  lies on this cubic; the edge  $x_2 = x_4 = 0$  cuts the cubic in the points  $O_1$  and  $(1, 0, 1, 0)$ . The edge  $x_1 = x_4 = 0$  cuts the cubic in the point  $O_2$  and  $(0, 1, 1, 0)$ . The point  $(1, 1, 1, 0)$  is therefore characterized.

The osculating cubic of the plane section of  $S$  by the plane  $\xi = 0$  has the equation

$$\xi = 0, \quad \eta^2 = \xi^3.$$

The projection of this cubic from the vertex  $(0, 0, 1, 0)$  is the cubic cone

$$x_2^2 x_4 = x_1^3.$$

The line joining  $O$  to  $(1, 1, 1, 0)$  intersects this cubic cone in the unit point of the tetrahedron of reference.

Hence the tetrahedron of reference giving rise to the expansion (22) and the unit point of the system have been geometrically characterized. Other properties of the surface could be obtained by considering the loci or envelopes of the various curves and surfaces associated with the sections of  $S$  through the arbitrary spinal tangent  $l$  in the plane  $\eta = 0$  by allowing  $l$  to vary in that plane. We shall refrain from discussing these details in this paper.

MICHIGAN STATE COLLEGE,  
EAST LANSING, MICH.

# CONDITIONS ON $u(x, y)$ AND $v(x, y)$ NECESSARY AND SUFFICIENT FOR THE REGULARITY OF $u+iv$ \*

BY

PHILIP T. MAKER

**Introduction.** On what subset of an open set must a continuous function of a complex variable be assumed to have a derivative in order that the regularity of the function be implied? A series of researches on this problem culminates in the following theorem due to Besicovitch:† A complex function  $f$ , continuous on an open set  $G$ , is regular in  $G$  if it is derivable at almost all the points of  $G$  and if further,

$$\limsup_{h \rightarrow 0} | [f(z+h) - f(z)]/h | < +\infty$$

at each point  $z$  of  $G$  except at most those of a set which is the sum of a sequence of sets of finite length.‡

At the same time the problem of reducing the conditions on  $u(x, y)$  and  $v(x, y)$ , where  $u+iv=f(z)$ , necessary and sufficient for the regularity of  $f(z)$ , has also received much attention,§ the most general result being the theorem of Looman and Menchoff:¶ If the functions  $u(x, y)$  and  $v(x, y)$ , continuous in an open set  $G$ , are derivable with respect to  $x$  and with respect to  $y$  at each point of  $G$  except at most at the points of an enumerable set, and if  $u_x(x, y) = v_y(x, y)$  and  $u_y(x, y) = -v_x(x, y)$  at almost all the points  $(x, y)$  of  $G$ , then the function  $u+iv$  is regular in  $G$ .

In the first part of this paper we investigate a question raised by Saks,¶ as to the existence of a more general theorem including these two results. The answer obtained is affirmative in the case when the sequence of sets of finite length mentioned in the theorem of Besicovitch is an  $F_\sigma$  with respect to the open set considered.

In the second part we further extend the set on which the partial deriva-

\* Presented to the Society, April 16, 1938; received by the editors March 14, 1938.

† Proceedings of the London Mathematical Society, vol. 32, pp. 1-9. The version given here is due to S. Saks, *The Theory of the Integral*, New York, 1937, p. 197.

‡ For a definition of the length of a planar set, see C. Carathéodory, *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen*, 1914, pp. 404-426.

§ For a review of recent results, see D. Menchoff, *Les Conditions de Monogénéité*, *Actualités Scientifiques et Industrielles*, no. 329, 1936.

¶ Quoted from Saks, p. 199.

¶ Saks, p. 201.

tives of  $u$  and  $v$  may fail to exist by placing certain restrictions on  $u$  and  $v$  in addition to continuity. In particular, we state a condition on  $u$  and  $v$  according to which if the Dini partial derivatives of  $u$  and  $v$  are finite except at most on an  $F_\sigma$  of measure zero and if the partials, where they exist, satisfy the Cauchy-Riemann equations ( $u_x = v_y$ ,  $u_y = -v_x$ ) a.e. (almost everywhere), then  $u + iv$  is regular.

Finally, in the third part we obtain conditions necessary and sufficient in order that a function of two real variables be harmonic.

We agree to the following conventions:  $G$  will be used to denote an open set in the complex plane;  $R$  an open rectangle with sides  $x = a_1$ ,  $x = a_2$ ,  $y = b_1$ ,  $y = b_2$ , ( $a_1 < a_2$ ,  $b_1 < b_2$ );  $(R)$  will denote the boundary of  $R$ ;  $R$  will be said to be in  $G$  if  $R + (R)$  is contained in  $G$ .

1. **Extension of the Looman-Menchoff theorem.** We prove the following theorem:

**THEOREM 1.** *Let  $\mathcal{F}$  be any class of continuous functions defined in  $G$ . Let  $E_n$ , ( $n = 1, 2, \dots$ ), be subsets of  $G$ , closed with respect to  $G$ , and such that regularity of any function of  $\mathcal{F}$  on  $G - E_n$  implies its regularity throughout  $G$ .*

*If  $f(z) = u(x, y) + iv(x, y)$ , belonging to  $\mathcal{F}$ , has its Dini derivatives infinite on  $\sum_{n=1}^{\infty} E_n$  at most, and if the partial derivatives, where they exist, satisfy the Cauchy-Riemann equations a.e.,  $f(z)$  is regular in  $G$ .*

**Proof.\*** Let  $F$  be the points of  $G$  where  $f(z)$  is not regular.  $F$  is evidently closed in  $G$ , and it has to be proved that  $F$  is empty.

Suppose therefore, if possible, that  $F \neq \emptyset$ , and let  $F_n$  denote, for each positive integer  $n$ , the set of points of  $G$  such that whenever  $|h| \leq 1/n$ , none of the four differences

$$\begin{array}{ll} u(x+h, y) - u(x, y), & v(x+h, y) - v(x, y), \\ u(x, y+h) - u(x, y), & v(x, y+h) - v(x, y) \end{array}$$

exceeds  $|nh|$  in absolute value. By continuity of  $u$  and  $v$  each set  $F_n$  is closed in  $G$ . Every point of  $G$ , except for the set  $\sum_{n=1}^{\infty} E_n$ , has all derivatives finite and so falls in some  $F_n$ . The subset  $F$  of  $G$  can therefore be expressed as

$$F = \sum_{n=1}^{\infty} F_n \cdot F + \sum_{n=1}^{\infty} E_n \cdot F.$$

Since  $F$  is closed with respect to  $G$ , we have by Baire's theorem† that some term in the right-hand side is everywhere dense with respect to a portion of

\* The first part of this proof follows closely the first part of the proof which Saks gives of the Looman-Menchoff theorem, p. 199.

† Saks, p. 54.

$F$ , say  $S \cdot F$  in the square  $S$ . We have two cases according as this is true of a term in the first or the second sum.

If  $F_n \cdot F$  is supposed everywhere dense with respect to  $S \cdot F$ , then, since  $F_n \cdot F$  is closed in  $G$  and so with respect to  $S$ ,  $F_n \cdot F$  contains  $S \cdot F$ . Hence each point of  $G \cdot S$  has finite derivatives; and by an extension\* of the Looman-Menchoff theorem,  $S \cdot F = 0$ .

In the second case  $E_n$  is everywhere dense with respect to  $S \cdot F$ . Since  $E_n$  is closed with respect to  $G$ , it contains  $S \cdot F$ . We have therefore the case of  $f(z)$  regular in  $G \cdot S - E_n$ , and by our assumption about the sets  $E_n$  it follows that  $f(z)$  is regular throughout  $G \cdot S$ , contrary to the assumption that  $S \cdot F \neq 0$ . This completes the proof of Theorem 1.

By the theorem of Besicovitch it follows that a continuous function is regular in  $G$  if it is regular except for a set of finite length closed with respect to  $G$ . As an immediate consequence of Theorem 1 we obtain the following generalization of the Looman-Menchoff theorem:

**THEOREM 2.** *If  $f(z) = u + iv$  is continuous in  $G$ , and if the partial derivatives of  $u$  and  $v$  are infinite at most on the sum of a sequence of sets of finite length closed in  $G$ , and if the Cauchy-Riemann equations hold a.e. where the partials exist,  $f(z)$  is regular in  $G$ .*

It might be supposed that if the partial derivatives of  $u$  and  $v$  are assumed finite except for a closed set of measure zero, then the continuous function  $u + iv$  is regular in  $G$ . The following example shows that this is not the case.

Let  $u(0, y)$ ,  $(0 \leq y \leq 1)$ , be the function defined by Hille and Tamarkin,† monotone and continuous but not absolutely continuous, and let  $u(x, y) = u(0, y)$ ,  $(0 < x \leq 1; 0 \leq y \leq 1)$ . Then  $u(x, y)$  is continuous (even of bounded variation, a property to be defined in §2) with  $u_x(x, y) = 0$  and  $u_y(x, y) = 0$  a.e. The function  $u + iv$  is regular a.e. but not regular throughout the unit square.

**THEOREM 3.** *There exists a function  $f(z)$  defined in the unit square, continuous and of bounded variation there, and regular everywhere except on a closed set of measure zero.*

The function  $f(z)$  just defined is not regular on a set of parallel lines which divide the region into an infinite number of separated sets. The following question then arises: If  $f(z)$  is continuous in  $R$  and regular a.e. with the points of regularity connected, is  $f(z)$  regular in  $R$ ? For example, take a Cantor set

\* Saks, p. 200: "instead of assuming partial derivability of the functions  $u$  and  $v$ , it is sufficient to suppose that at each point of  $G$  (except at most those of an enumerable set) these functions have with respect to each variable,  $x$  and  $y$ , their partial Dini derivatives finite."

† American Mathematical Monthly, vol. 36 (1929), pp. 255-264.

on two adjacent sides of  $R$  and pass parallels to the sides through these sets, and take the intersection points of these lines as the set  $E$ . If  $f(z)$  is continuous in  $R$  and regular in  $R-E$ , is  $f(z)$  regular on  $E$  as well? In the following section we shall show that this is the case under the added assumption that  $u$  and  $v$  are of bounded variation.

**2. Functions with summable partials.** In this section we assume explicitly or otherwise, that the partial derivatives  $u_x, u_y, v_x, v_y$  exist a.e. and are summable as functions of two variables. We begin by considering a function  $f(z)$  defined only on a certain subset of  $G$  and state conditions under which  $f(z)$  is equal there to a function regular throughout  $G$ .

We need the following definitions: The intersection of an open set  $G$  and any set of almost all lines parallel to the  $x$  ( $y$ ) axis will be called a  $G_x$  ( $G_y$ ). We define  $R_x$  and  $R_y$  similarly. The sum of a  $G_x$  and a  $G_y$  will be denoted by  $G_x + G_y$ .

To avoid the phrase "continuity of  $u(x, y)$  as a function of  $x$  for almost all values of  $y$ ," we shall say " $u(x, y)$  is continuous in  $x$  for almost all  $y$ ," or more simply, " $u(x, y)$  is continuous in a  $G_x$ ."

**THEOREM 4.** Let  $f(z)$  be defined in a  $G_x + G_y$  with  $f(z) = u(x, y) + iv(x, y)$ . Suppose that in every  $R$  in  $G$  the following conditions hold:

(a)  $u(x, y)$  and  $v(x, y)$  are absolutely continuous in  $x$  for almost all  $y$ , and in  $y$  for almost all  $x$ .

(b) The partials  $u_x, u_y, v_x, v_y$ , are Lebesgue-summable in  $(x, y)$ .

(c) The Cauchy-Riemann equations hold a.e.

Then  $f(z)$  is equal in  $G_x + G_y$  to a function regular in  $G$ .

**Proof.** Let  $P$  be any point of  $G$ , and let  $R$  be a rectangle in  $G$  containing  $P$ , with the vertex  $(a_1, b_1)$  chosen so that  $u(x, b_1)$  and  $v(x, b_1)$  are continuous in  $x$  for  $a_1 \leq x \leq a_2$ , and  $u(a_1, y)$  and  $v(a_1, y)$  are continuous in  $y$  for  $b_1 \leq y \leq b_2$ . By Fubini's theorem and the Cauchy-Riemann relation we have

$$\int_{a_1}^x \int_{b_1}^y u_y(\xi, \eta) d\eta d\xi = \int_{b_1}^y \int_{a_1}^x u_x(\xi, \eta) d\xi d\eta = - \int_{b_1}^y \int_{a_1}^x v_x(\xi, \eta) d\xi d\eta, \quad (x, y) \text{ in } R.$$

By the absolute continuity of  $u(x, y)$  in  $R_y$  we have

$$\int_{b_1}^y u_x(\xi, \eta) d\eta = u(\xi, y) - u(\xi, b_1) \text{ in } R_y.$$

Similarly

$$\int_{a_1}^x v_x(\xi, \eta) d\xi = v(x, \eta) - v(a_1, \eta) \text{ in some } R_x.$$

Substituting these in the extreme members of the first equation, we have the relation

$$\int_{a_1}^x [u(\xi, y) - u(\xi, b_1)] d\xi = - \int_{b_1}^y [v(x, \eta) - v(a_1, \eta)] d\eta, \quad (x, y) \text{ in } R.$$

Let

$$\begin{aligned} w_1(x, y) &= \int_{a_1}^x u(\xi, y) d\xi - \int_{b_1}^y v(a_1, \eta) d\eta \\ &= - \int_{b_1}^y v(x, \eta) d\eta + \int_{a_1}^x u(\xi, b_1) d\xi. \end{aligned}$$

By the continuity of  $u$  in  $R_x$  and the first equality,  $w_{1x}(x, y) = u(x, y)$  in  $R_x$ . By the continuity of  $v$  in  $R_y$  and the second equality,  $w_{1y}(x, y) = -v(x, y)$  in  $R_y$ .

Again,

$$\int_{b_1}^y \int_{a_1}^x u_x(\xi, \eta) d\xi d\eta = \int_{a_1}^x \int_{b_1}^y u_x(\xi, \eta) d\eta d\xi = \int_{a_1}^x \int_{b_1}^y v_y(\xi, \eta) d\eta d\xi,$$

and as before

$$\int_{b_1}^y [u(x, \eta) - u(a_1, \eta)] d\eta = \int_{a_1}^x [v(\xi, y) - v(\xi, b_1)] d\xi, \quad (x, y) \text{ in } R.$$

Let

$$\begin{aligned} w_2(x, y) &= \int_{b_1}^y u(x, \eta) d\eta + \int_{a_1}^x v(\xi, b_1) d\xi \\ &= \int_{a_1}^x v(\xi, y) d\xi + \int_{b_1}^y u(a_1, \eta) d\eta. \end{aligned}$$

By the continuity of  $u$  in  $R_y$  and the first equality,  $w_{2y}(x, y) = u(x, y)$  in  $R_y$ . By the continuity of  $v$  in  $R_x$  and the second equality,  $w_{2x}(x, y) = v(x, y)$  in  $R_x$ .

By a theorem due to H. Rademacher,\* a function  $\phi(x, y) + i\psi(x, y)$  is regular in  $R$  when the following conditions are satisfied there: (1)  $\phi(x, y)$  and  $\psi(x, y)$  are absolutely continuous in  $x$  and  $y$  separately for all values of  $x$  and  $y$ ; (2)  $\phi(x, y)$  and  $\psi(x, y)$  are summable in  $(x, y)$ ; (3)  $\phi_x, \phi_y, \psi_x,$  and  $\psi_y$ , which necessarily exist a.e., are summable in  $(x, y)$ ; (4) the Cauchy-Riemann equations hold a.e. This result is applicable to  $w_1 + iw_2$ . For (1) is true since by definition  $w_1$  and  $w_2$  are integrals; (4) has already been verified; (3) is true since the summability of  $u$  and  $v$  is a consequence of the summability of the partials and the continuity of  $u(a_1, y)$ , as the following inequality shows:

\* Mathematische Zeitschrift, vol. 4 (1919), p. 184, Theorem II.

$$\begin{aligned}
\int_{a_1}^{a_2} \int_{b_1}^{b_2} |u| dy dx &= \int_{a_1}^{a_2} \int_{b_1}^{b_2} \left| \int_{a_1}^x u_z(\xi, y) d\xi \right| dy dx \\
&+ \int_{a_1}^{a_2} \int_{b_1}^{b_2} |u(a_1, y)| dy dx \\
&\leq \int_{a_1}^{a_2} \int_{b_1}^{b_2} \int_{a_1}^{a_2} |u_z(\xi, y)| d\xi dy dx \\
&+ \int_{a_1}^{a_2} \int_{b_1}^{b_2} |u(a_1, y)| dy dx.
\end{aligned}$$

By repeating this argument for  $w_1$  and  $w_2$ , the partials of which are summable by (3), we obtain (2). Hence  $w_1 + iw_2$  is regular in  $R$ .

If  $R_1$  and  $R_2$  are any two separated rectangles lying in a domain of  $G$ , the regular function to which  $u + iv$  is equal in  $R_1$  is readily shown to be the analytic continuation of the regular function to which  $u + iv$  is equal in  $R_2$ ; hence the proof is complete.

**COROLLARY.** *If, in addition to satisfying the conditions of Theorem 4,  $u(x, y)$  is continuous in  $x$  or in  $y$ , and  $v(x, y)$  is continuous in  $x$  or in  $y$ , throughout  $G$ , then  $u + iv$  is regular in  $G$ .*

For, in  $G_x + G_y$ ,  $u + iv$  is equal to a regular function. On any line parallel to the  $x$  axis, for example, the points of  $G_y$  are everywhere dense. Hence the continuity of  $u(x, y)$  with respect to  $x$  implies that  $u(x, y)$  is equal everywhere in  $G$  to the real part of a regular function, and by a similar argument applied to  $v(x, y)$ , the proof is complete.

A function  $f(x)$  satisfies the (N) condition in the set  $A$  if the measure of the image of  $E$  ( $E \subset A$ ) with respect to  $f(x)$  is zero when the measure of  $E$  is zero. A function  $f(x, y)$  will be said to *satisfy the (N) condition linearly in  $A$* , a plane set, if for almost all  $x_0$  and  $y_0$  in the interval  $(-\infty, +\infty)$ ,  $f(x_0, y)$  and  $f(x, y_0)$  satisfy the (N) condition in  $A$ .

**THEOREM 5.** *Theorem 4 holds when condition (a) is replaced by either of the following:*

- (i)  $u_x$  and  $v_x$  exist and are finite in a  $G_x$ ;  $u_y$  and  $v_y$  exist and are finite in a  $G_y$ .
- (ii)  $u(x, y)$  and  $v(x, y)$  are continuous in  $x$  in a  $G_x$ , in  $y$  in a  $G_y$ , and satisfy the (N) condition in a  $G_x$  and a  $G_y$ .

**Proof.** It is readily seen from the proof of Theorem 4, and the similarity of the conditions on  $u$  and  $v$  in  $x$  and in  $y$ , that it will suffice to show that (i) and (ii) imply the absolute continuity of  $u$  with respect to  $x$  in an  $R_x$  for all  $R$  of  $G$ .

If (i) is given, from the summability in  $(x, y)$  of  $u_x$  it follows that  $u_x$  is summable in  $x$  for almost all  $y$ . Hence for almost all  $y$ ,

$$\int_{a_1}^x u_x(\xi, y) d\xi = u(x, y) - u(a_1, y)$$

by a well known theorem,\* so  $u(x, y)$  is absolutely continuous in  $x$  in a  $G_x$ .

To prove that (ii) implies the absolute continuity of  $u(x, y)$  in  $x$  in  $R_x$ , let  $E_y$  denote the linear set in  $R$  with ordinate  $y$  at which  $u(x, y)$  has a finite nonnegative partial with respect to  $x$ , and let  $E$  be all such points in  $R$ ; then since  $u_x(x, y)$  is summable on  $E$ ,  $\int_{E_y} u_x(\xi, y) d\xi < +\infty$  for almost all  $y$ , ( $b_1 \leq y \leq b_2$ ), and by a theorem due to N. Bary† this is sufficient for the absolute continuity of  $u(x, y)$  in  $x$  in  $G_x$ , assumed continuous and satisfying the (N) condition for almost all  $y$ .

Given  $f(x, y)$  continuous in  $\bar{R}$ , denote by  $T_1(f; x; b_1, b_2)$ , for any  $x$  such that  $a_1 \leq x \leq a_2$ , the total variation of  $f(x, y)$  with respect to  $y$  from  $b_1$  to  $b_2$ , and by  $T_2(f; y; a_1, a_2)$ , for any  $y$  such that  $b_1 \leq y \leq b_2$ , the total variation of  $f(x, y)$  with respect to  $x$  from  $a_1$  to  $a_2$ . When  $T_1$  and  $T_2$  are finite for almost all  $x$  and  $y$ , respectively, and the Lebesgue integrals  $\int_{a_1}^{a_2} T_1(f; x; b_1, b_2) dx$  and  $\int_{b_1}^{b_2} T_2(f; y; a_1, a_2) dy$  exist (finite),  $f(x, y)$  is said to be of *bounded variation* (in the sense of Tonelli).‡

**LEMMA.** *If the continuous function  $f(x, y)$  (a) is absolutely continuous in  $x$  in  $R_x$  and in  $y$  in  $R_y$ , or (b) has  $f_x$  existing and finite in  $R_x$  and  $f_y$  existing and finite in  $R_y$ , a necessary and sufficient condition that  $f(x, y)$  be of bounded variation in  $R$  is that  $f_x$  and  $f_y$  be Lebesgue-summable there.*

**Proof.** To prove the necessity, assume that both  $\int_{a_1}^{a_2} T_1(f; x; b_1, b_2) dx$  and  $\int_{b_1}^{b_2} T_2(f; y; a_1, a_2) dy$  exist and are finite. Since by either (a) or (b)  $f(x, y)$  is absolutely continuous in  $x$  in some  $R_x$  and in  $y$  in some  $R_y$ ,

$$T_1(f; x; b_1, b_2) = \int_{b_1}^{b_2} |f_y| dy, \quad T_2(f; y; a_1, a_2) = \int_{a_1}^{a_2} |f_x| dx$$

for almost all  $x$  and  $y$ , respectively.§ Substituting these values for  $T_1$  and  $T_2$ , we get  $\int_{a_1}^{a_2} \int_{b_1}^{b_2} |f_y| dy dx$  and  $\int_{b_1}^{b_2} \int_{a_1}^{a_2} |f_x| dx dy$  finite; hence the double integrals exist.||

The converse follows from the fact that the Lebesgue integral is absolutely

\* E. W. Hobson, *The Theory of Functions of a Real Variable*, vol. 1, 3d edition, Cambridge, 1927, p. 601.

† Saks, p. 285.

‡ Saks, p. 169.

§ Hobson, p. 605.

|| Hobson, p. 631.

convergent. For example, the summability of  $f_y$  implies the finite existence of

$$\int_{a_1}^{a_2} T_1(f; x; b_1, b_2) dx = \int_{a_1}^{a_2} \int_{b_1}^{b_2} |f_y| dy dx.$$

As a consequence of this lemma and Theorem 4, it follows that if  $u$  and  $v$  are of bounded variation in every  $R$  of  $G$  and absolutely continuous in  $x$  in  $G_x$  and in  $y$  in  $G_y$ , and if the Cauchy-Riemann equations are satisfied a.e. in  $G$ ,  $f(z)$  is regular in  $G$ . Using the following definition, we can state Theorem immediately.

The function  $f(x, y)$  is *absolutely continuous* in  $R$  if it is of bounded variation in  $(x, y)$  and absolutely continuous in  $x$  in  $R_x$  and in  $y$  in  $R_y$ .\*

**THEOREM 6.** *The function  $u+iv$  is regular in  $G$  if and only if  $u$  and  $v$  are absolutely continuous in every  $R$  of  $G$  and if the Cauchy-Riemann equations hold a.e. in  $G$ .*

**THEOREM 7.** *If  $u$  and  $v$  are of bounded variation in every  $R$  of  $G$ , and if their partial derivatives are infinite at most on an  $F_\sigma$  with respect to  $G$ , of measure zero, on which  $u$  and  $v$  satisfy the (N) condition linearly, and if the Cauchy-Riemann equations hold a.e. where the partials exist,  $u+iv$  is regular in  $G$ .*

**Proof.** We first show that if  $u$  and  $v$  are of bounded variation in  $R$ , and  $u+iv$  is regular except possibly for a set  $E$  of measure zero, necessarily closed with respect to  $R$ , on which  $u$  and  $v$  satisfy the (N) condition, then  $u+iv$  is regular in  $R$ . By Theorem 6, it will be sufficient to show that  $u$  and  $v$  are absolutely continuous in  $R$ ; we need therefore to prove that  $u$  and  $v$  are absolutely continuous in  $x$  in an  $R_x$  and in  $y$  in an  $R_y$ . By the symmetry of the conditions of  $u$  and  $v$  with respect to  $x$  and  $y$ , it will suffice to prove that  $u$  is absolutely continuous in  $x$  in an  $R_x$ . Any line  $L \equiv [y = y_0]$  in  $R$  is composed of  $\sum_{n=1}^{\infty} I_n + E \cdot L$ , where  $I_n$ , ( $n = 1, 2, \dots$ ), is an interval free of  $E$  and like  $E \cdot L$  may be null. Since  $u(x, y_0)$  satisfies the (N) condition on each of these subsets, it clearly satisfies the condition on their sum, that is, on  $L$ . Hence, because  $u$  is of bounded variation and satisfies the (N) condition in an  $R_x$ ,  $u$  is absolutely continuous in  $x$  in an  $R_x$ .†

Now use Theorem 1 and the proof is complete.

In the example given in the first section,  $u$  and  $v$  are of bounded variation, and their partials exist a.e. satisfying the Cauchy-Riemann equations; hence the assumption of the (N) condition or its equivalent must necessarily be made in this theorem. However there is a less general class of sets of measure zero for which this (N) condition obviously need not be assumed.

\* Saks, p. 169.

† Saks, p. 227.

**COROLLARY.** *If  $u(x, y)$  and  $v(x, y)$  are of bounded variation in every  $R$  of  $G$ , and if their partial derivatives are infinite at most on an  $F_\sigma$  with respect to  $G$  such that the lines parallel to the axes which intersect  $F_\sigma$  in a nondenumerable number of points are of measure zero, and if the Cauchy-Riemann equations hold a.e. where the partials exist, then  $u+iv$  is regular in  $G$ .*

**COROLLARY.** *Let  $E$  be a non-empty subset of  $G$  such that almost every line parallel to an axis intersects  $E$  in at most a denumerable set. The function  $f(z)$ , regular in  $G-E$ , is regular throughout  $G$  if and only if  $u$  and  $v$  are of bounded variation in every  $R$  of  $G$ .*

This corollary should be compared with a similar result of V. Fedoroff.\* He requires that  $G-E$  be connected; otherwise  $E$  is any set of measure zero. On the other hand, his condition (D) is more restrictive than the present condition that  $u$  and  $v$  be of bounded variation.

**3. Morera's theorem and its application to harmonic functions.** Rademacher† has shown that if for every  $R$  in  $G$ ,  $u$  and  $v$  are summable in  $(x, y)$  and, as functions of  $x$  and  $y$  separately, are summable for each value of  $x$  and  $y$ , and if  $\int_{(R)} (u+iv) dz = 0$ , then  $u+iv$  is regular in  $G$  except for removable discontinuities of measure zero.

In extending this result we introduce the following definition: Let  $E$  be a set of measure zero in  $G$ , and let  $R$  be any rectangle in  $G$  with sides  $x=a_1$ ,  $x=a_2$ ,  $y=b_1$ ,  $y=b_2$ , ( $a_1 < a_2$ ,  $b_1 < b_2$ ), with  $(a_1, b_1)$  and  $(a_2, b_2)$  in  $G-E$ ; the set of all such rectangles will be called *almost all  $R$  in  $G$* .

**THEOREM 8.** *If for almost all  $R$  in  $G$  the functions  $u$  and  $v$  are summable in  $(x, y)$  and  $\int_{(R)} (u+iv) dz = 0$ ,  $u+iv$  is regular in  $G$  except possibly for removable discontinuities of measure zero.*

**Proof.** Let

$$(1) \quad \begin{aligned} w_1(x, y) = & - \int_{a_1}^x \int_{b_1}^y v(\xi, \eta) d\eta d\xi + \int_{a_1}^x \int_{a_1}^{\xi} u(\xi_1, b_1) d\xi_1 d\xi \\ & - \int_{b_1}^y \int_{b_1}^{\eta} u(a_1, \eta_1) d\eta_1 d\eta, \end{aligned}$$

$$(2) \quad \begin{aligned} w_2(x, y) = & \int_{a_1}^x \int_{b_1}^y u(\xi, \eta) d\eta d\xi + \int_{a_1}^x \int_{a_1}^{\xi} v(\xi_1, b_1) d\xi_1 d\xi \\ & - \int_{b_1}^y \int_{b_1}^{\eta} u(a_1, \eta_1) d\eta_1 d\eta. \end{aligned}$$

\* Recueil Mathématique de la Société de Moscou, vol. 41 (1934), p. 97.

† Rademacher, loc. cit., p. 183, Theorem I.

The condition  $\int_{(R)} (u + iv) dz = 0$  yields

$$(3) \quad \int_{a_1}^x u(\xi, b_1) d\xi - \int_{a_1}^x u(\xi, y) d\xi - \int_{b_1}^y v(x, \eta) d\eta + \int_{b_1}^y v(a_1, \eta) d\eta = 0,$$

$$(4) \quad \int_{a_1}^x v(\xi, b_1) d\xi - \int_{a_1}^x v(\xi, y) d\xi + \int_{b_1}^y u(x, \eta) d\eta - \int_{b_1}^y u(a_1, \eta) d\eta = 0.$$

In the first term of  $w_1(x, y)$  replace  $\int_{b_1}^y v(\xi, \eta) d\eta$  by the function

$$\int_{a_1}^x u(\xi, b_1) d\xi - \int_{a_1}^x u(\xi, y) d\xi + \int_{b_1}^y v(a_1, \eta) d\eta$$

to which by (3) it is equal a.e. Since the value of  $w_1(x, y)$  has not altered and the new integrand is continuous in  $x$ , the latter equals the partial,  $w_{1x}(x, y)$  in an  $R_x$ .

Hence by (3),

$$(5) \quad w_{1x}(x, y) = - \int_{b_1}^y v(x, \eta) d\eta + \int_{a_1}^x u(\xi, b_1) d\xi \quad \text{a.e.}$$

Replacing  $\int_{a_1}^x v(\xi, \eta) d\xi$  in the first term of  $w_1(x, y)$  by

$$\int_{a_1}^x v(\xi, b_1) d\xi + \int_{b_1}^y u(x, \eta) d\eta - \int_{b_1}^y u(a_1, \eta) d\eta$$

and using a similar argument, we have

$$(6) \quad w_{1y}(x, y) = - \int_{a_1}^x v(\xi, y) d\xi - \int_{b_1}^y u(a_1, \eta) d\eta \quad \text{a.e.}$$

Applying a similar proof to  $w_2(x, y)$ , we obtain

$$(7) \quad w_{2x}(x, y) = \int_{b_1}^y u(x, \eta) d\eta + \int_{a_1}^x v(\xi, b_1) d\xi \quad \text{a.e.},$$

$$(8) \quad w_{2y}(x, y) = \int_{a_1}^x u(\xi, y) d\xi - \int_{b_1}^y v(a_1, \eta) d\eta \quad \text{a.e.}$$

By (3), (5), and (8), we verify  $w_{1x} = w_{2y}$  a.e.; and by (4), (6), and (7),  $w_{1y} = -w_{2x}$  a.e. Theorem 4 implies that  $w_1 + iw_2$  is regular in  $R$ , therefore, and because  $u + iv$  is equal a.e. to the second derivative of  $w_1 + iw_2$ , the theorem is proved.

As an application of Theorem 8, consider a function  $f(x, y)$  of two real variables, with partials  $f_x$  and  $f_y$ . If we take these to be the functions  $u$  and  $v$ , formally the conditions  $\int_{(R)} f_x dx + f_y dy = 0$  and  $\int_{(R)} (df/dn) ds = 0$  are the real

and imaginary parts, respectively, of  $\int_{(R)}(u+iv)dz=0$ . Consequently,  $u+iv$  is regular and  $f(x, y)$ , harmonic, except for removable discontinuities. A precise statement of the conditions for this conclusion is the following:

**THEOREM 9.** *A necessary and sufficient condition that  $f(x, y)$  be equal in a  $G_x+G_y$  to a function harmonic in  $G$  is that for almost all  $R$  in  $G$  the following conditions hold:*

- (i)  $f(x, y)$  is absolutely continuous in  $x$  in an  $R_x$ , and in  $y$  in an  $R_y$ .
- (ii)  $f_x(x, y)$  and  $f_y(x, y)$  are summable in  $R$ .
- (iii)  $\int_{(R)}(df/dn)ds=0$ .

As in the case of Theorem 4, the requirement of absolute continuity in  $x$  in  $R_x$  and in  $y$  in  $R_y$  can be replaced by conditions (i) or (ii) of Theorem 5.

The further assumption that  $f(x, y)$  is continuous in  $x$  or in  $y$  separately throughout  $G$  would imply that  $f$  is harmonic everywhere in  $G$ .

Finally, in view of the lemma, we state the analogue of Theorem 6:

**THEOREM 10.** *The function  $f(x, y)$  is harmonic in  $G$  if and only if for almost all  $R$  in  $G$ ,  $f(x, y)$  is absolutely continuous in  $R$  and  $\int_{(R)}(df/dn)ds=0$ .*

Theorems 9 and 10 are similar to results obtained by G. C. Evans\* in connection with the theory of potential functions.

---

\* G. C. Evans, *Fundamental points of potential theory*, The Rice Institute Pamphlet, vol. 7 no. 4, 1920, p. 286.

# IRREDUCIBLE SYSTEMS OF ALGEBRAIC DIFFERENTIAL EQUATIONS\*

BY

WALTER CHARLES STRODT

## INTRODUCTION

Let  $\mathcal{Y}$  be a domain of rationality, and let  $y_1, \dots, y_n$  be a set of indeterminates. Then the set of prime ideals in the ring of polynomials  $\mathcal{Y}[y_1, \dots, y_n]$  satisfies a divisor-chain condition for decreasing sequences as well as for increasing sequences. That is, a sequence of prime ideals  $\Sigma_1, \Sigma_2, \dots$  in  $\mathcal{Y}[y_1, \dots, y_n]$  must be of finite length not only if  $\Sigma_{i+1}$  properly includes  $\Sigma_i$  for every  $i$ , but also if  $\Sigma_{i+1}$  is properly included in  $\Sigma_i$  for every  $i$ .†

However, if the domain of rationality  $\mathcal{Y}$  is a set of functions meromorphic in an open region  $\mathcal{A}$ , and if  $\mathcal{Y}$  is closed to differentiation (in other words, if  $\mathcal{Y}$  is a *field*, in the terminology of algebraic differential equations‡), and if  $\mathcal{Y}\{y_1, \dots, y_n\}$  is the differential-ring consisting of all forms in  $y_1, \dots, y_n$  with coefficients in  $\mathcal{Y}$ , then the set of prime differential-ideals in  $\mathcal{Y}\{y_1, \dots, y_n\}$  satisfies a divisor-chain condition for increasing sequences,§ but does not satisfy such a condition for decreasing sequences. That is, we can have an infinite sequence  $\Sigma_1, \Sigma_2, \dots$  of prime differential-ideals such that  $\Sigma_{i+1}$  is properly included in  $\Sigma_i$ . In the set-theoretic sense the sequence  $\Sigma_1, \Sigma_2, \dots$  converges to a limiting set  $\Sigma$  which is the intersection of the  $\Sigma_i$ . If  $\mathcal{M}_i$  is the manifold of  $\Sigma_i$ , then the sequence  $\mathcal{M}_1, \mathcal{M}_2, \dots$  is a monotonically increasing sequence converging in the set-theoretic sense to a set  $\mathcal{N}$  which is the union of the  $\mathcal{M}_i$ . However, while the limiting set  $\Sigma$  is a prime differential-ideal, the limiting set  $\mathcal{N}$  not only is not the manifold of  $\Sigma$ , but is not a manifold at all. We are concerned in this paper with the relation between  $\mathcal{N}$  and the manifold  $\mathcal{M}$  of  $\Sigma$ .

In the terminology of ADE, what we are considering is an infinite sequence  $\Sigma_1, \Sigma_2, \dots$  of closed irreducible systems in  $y_1, \dots, y_n$  such that  $\mathcal{M}_i$ , the manifold of  $\Sigma_i$ , is a proper part of the manifold of  $\Sigma_{i+1}$ , ( $i = 1, 2, \dots$ ).

\* Presented to the Society, February 26, 1938, under the title *Sequences of systems of algebraic differential equations*; received by the editors April 12, 1938.

† Cf. Van der Waerden, *Moderne Algebra*, vol. 2, pp. 25, 63. The set of all ideals in  $\mathcal{Y}[y_1, \dots, y_n]$  satisfies a divisor-chain condition for increasing sequences, but not for decreasing sequences.

‡ See, for example, J. F. Ritt, *Differential Equations from the Algebraic Standpoint*, American Mathematical Society Colloquium Publications, vol. 14, New York, 1932. We shall refer to this book by the letters ADE, and we shall use the terminology of ADE without further reference.

§ As a consequence of Ritt's theorem on the completeness of infinite systems, ADE, §7. The set of all differential-ideals in  $\mathcal{Y}\{y_1, \dots, y_n\}$  does not satisfy a divisor-chain condition for increasing sequences (by ADE, §11).

The set  $\mathcal{N}$  is the union  $\mathcal{M}_1 + \mathcal{M}_2 + \dots$ , and  $\mathcal{M}$  is the manifold of the system  $\Sigma$  consisting of all forms  $F$  such that  $F$  is in every  $\Sigma_i$ .<sup>\*</sup> Not only does  $\mathcal{M}$  contain solutions which do not appear in  $\mathcal{N}$ , but there is even a sense in which we may say that  $\mathcal{M}$  is of *higher dimensionality* than  $\mathcal{N}$ . This is expressed in the statement that  $\Sigma$  has more arbitrary unknowns than  $\Sigma_i$ , ( $i=1, 2, \dots$ ) (Theorem 3). On the other hand, we shall see that  $\mathcal{M}$  may be described as the set of all ordered sets of  $n$  analytic functions which can be approximated in a certain manner by solutions in  $\mathcal{N}$  (Theorem 4).

Approximability, as we shall define it, will not imply the familiar uniform approximability in a region. Indeed, for certain sequences  $\Sigma_1, \Sigma_2, \dots$ , every solution of  $\Sigma$  which is not in  $\mathcal{N}$  possesses no region of analyticity in which it may be uniformly approximated by solutions of the  $\Sigma_i$ . On the other hand, there exist sequences  $\Sigma_1, \Sigma_2, \dots$  such that every solution of  $\Sigma$  has a region of analyticity in which it can be approximated uniformly by solutions of the  $\Sigma_i$ .

As a converse to Theorem 3, we have the theorem that for every closed irreducible system  $\Sigma$  with a non-empty set of arbitrary unknowns there is a sequence  $\Sigma_1, \Sigma_2, \dots$  of closed irreducible systems such that  $\Sigma_{i+1}$  holds  $\Sigma_i$ ,  $\Sigma$  is the set of forms common to the  $\Sigma_i$ , and  $\Sigma_i$  has fewer arbitrary unknowns than  $\Sigma$ , ( $i=1, 2, \dots$ ). In fact,  $\Sigma_i$  may be taken to have no arbitrary unknowns.

In studying the sequences  $\Sigma_1, \Sigma_2, \dots$  we use several preliminary theorems which are demonstrated in Part I of this paper. These theorems are extensions of results obtained by Ritt. Theorem 1 deals with the possibility of approximating a solution of a prime algebraic system by solutions which do not annul a specified simple form. Theorem 2 has to do with an analogous question for differential equations.

Lemmas 1 and 2 of Part II are devoted to the study of the degree of freedom which one enjoys in assigning initial conditions to a solution of a prime algebraic system.

#### PART I. APPROXIMATION THEOREMS

1. The following theorem is due to Ritt: *Let  $\Sigma$  be an indecomposable system of simple forms in  $y_1, \dots, y_n$ . Let  $B$  be any simple form which does not hold  $\Sigma$ . Given any solution of  $\Sigma$ , analytic in an open region  $\mathfrak{A}_1$ , there is an open region  $\mathfrak{A}'$ , contained in  $\mathfrak{A}_1$ , in which the given solution can be approximated uniformly, with arbitrary closeness, by solutions of  $\Sigma$  for which  $B$  is distinct from 0 throughout  $\mathfrak{A}'$ .*<sup>†</sup>

<sup>\*</sup> We shall see that  $\Sigma$  is closed and irreducible (Theorem 3, below).

<sup>†</sup> ADE, §64.

We shall use the following modification of Ritt's result:

**THEOREM 1.** *Let  $\Sigma$  be an indecomposable system of simple forms in  $y_1, \dots, y_n$ . Let  $B$  be any simple form which does not hold  $\Sigma$ . Given any solution of  $\Sigma$ , analytic in an open region  $\mathfrak{A}_1$ , there is an open region  $\mathfrak{M}$  in  $\mathfrak{A}_1$ , such that  $\mathfrak{A}_1 - \mathfrak{M}$  is isolated in  $\mathfrak{A}_1$ ,\* and such that for every bounded simply-connected open region  $\mathfrak{C}$  which lies with its boundary in  $\mathfrak{M}$ , there exists a sequence of solutions of  $\Sigma$ , analytic in  $\mathfrak{C}$ , for each of which  $B$  is distinct from zero throughout  $\mathfrak{C}$ , the sequence converging to the given solution in  $\mathfrak{C}$ , uniformly in every closed subset of  $\mathfrak{C}$ .*

Following the procedure in ADE, §64, we introduce the system  $\Sigma_1$  in  $z_1, \dots, z_n$  with which are associated the simple forms  $R, G, D, E_{ij}$ , ( $i = q+1, \dots, n; j = 0, 1, \dots, g-1$ ), as in §§59-61. We denote by  $C_i$  some simple form of  $\Sigma_1$  which is of degree  $m$  in the  $z_j$ , ( $j = 1, 2, \dots, q, q+i$ ), and of degree  $m$  in  $z_{q+i}$ , the coefficient of  $(z_{q+i})^m$  being unity ( $i = 1, \dots, p$ ). With  $\Sigma_1$  and  $B$  are associated the simple forms  $C, N, X$ , and  $Y$  of §64. We let  $H = XR + YN$ , and we may and do assume that  $X$  and  $Y$  are so chosen that  $H$  is divisible by  $DG$ . By  $\xi_1, \dots, \xi_n$  we understand that solution of  $\Sigma_1$  which corresponds, under the transformation of §57, to the given solution of  $\Sigma$ . Proceeding from this definition of  $H$  and the  $\xi_i$  as in §63, we introduce constants  $b_1, \dots, b_q$  such that  $H$ , under the substitution  $z_i = \xi_i + b_i$ , ( $i = 1, \dots, q$ ), becomes a function of  $x$  not identically zero. Then  $H$ , under the substitution  $z_i = \xi_i + b_i h$ , ( $i = 1, \dots, q$ ), becomes a polynomial

$$\alpha_r h^r + \alpha_{r+1} h^{r+1} + \dots + \alpha_s h^s,$$

where  $r \geq 1$ , the  $\alpha_j$  are functions of  $x$ , meromorphic in  $\mathfrak{A}_1$ , and  $\alpha_r(x) \neq 0$ .

Let  $\Pi$  be the set of simple forms  $C, N, X, Y, R, G, D, E_{ij}, C_1, \dots, C_p$ . Let  $\mathfrak{M}$  be the set of points of  $\mathfrak{A}_1$  at which the coefficients in  $\Pi$  are analytic and at which the function  $\alpha_r$  is different from zero. Evidently  $\mathfrak{A}_1 - \mathfrak{M}$  is isolated in  $\mathfrak{A}_1$ . The functions  $\alpha_j$  are analytic in  $\mathfrak{M}$ .

Let  $\mathfrak{C}$  be a bounded simply-connected open region which lies with its boundary in  $\mathfrak{M}$ . Since  $\mathfrak{C}$  is at a positive distance from the boundary of  $\mathfrak{M}$ , the function  $\alpha_r$  is bounded away from zero in  $\mathfrak{C}$ , and the functions  $\alpha_{r+1}, \dots, \alpha_s$  are bounded in  $\mathfrak{C}$ . This implies that for every sufficiently small nonzero constant  $h$  the polynomial  $\alpha_r h^r + \dots + \alpha_s h^s$  vanishes nowhere in  $\mathfrak{C}$ . Therefore in the considerations of ADE, §63, we may take  $\mathfrak{A}_2 = \mathfrak{C}$ . Moreover, we may take  $\mathfrak{A}_3 = \mathfrak{C}$ , since the functions  $\xi_1, \dots, \xi_q$  and the coefficients in  $C_1, \dots, C_p$  are bounded in  $\mathfrak{C}$ .

\* We shall say that a subset  $\mathfrak{S}$  of an open region  $\mathfrak{R}$  is isolated in  $\mathfrak{R}$  if  $\mathfrak{S}$  is empty, or if  $\mathfrak{S}$  is a non-empty set which has no limit points in  $\mathfrak{R}$ .

Let  $\mathfrak{D}$  be a region which lies with its boundary in  $\mathfrak{C}$ . Taking  $\mathfrak{A}' = \mathfrak{D}$ , and following the procedure of ADE, §63, we determine a sequence of solutions

$$\zeta_{1,i}, \dots, \zeta_{n,i}, \quad i = 1, 2, \dots,$$

of  $\Sigma_1$  for each of which  $H$  is distinct from zero throughout  $\mathfrak{C}$ , the sequence converging to  $\xi_1, \dots, \xi_n$  uniformly in  $\mathfrak{D}$ . Moreover, the sequence is so constructed that there is a positive number  $d'$  for which the inequalities

$$|\zeta_{ji}| < d', \quad j = 1, \dots, n; i = 1, 2, \dots,$$

are valid in  $\mathfrak{C}$ . Hence by Vitali's theorem\* the sequence  $\zeta_{1,i}, \dots, \zeta_{n,i}$  converges to  $\xi_1, \dots, \xi_n$  in  $\mathfrak{C}$ , uniformly in every closed subset of  $\mathfrak{C}$ . Corresponding to this sequence is a sequence of solutions of  $\Sigma$  for each of which  $B$  is distinct from zero throughout  $\mathfrak{C}$ , the sequence converging to the given solution in  $\mathfrak{C}$ , uniformly in every closed subset of  $\mathfrak{C}$ .

2. We use Theorem 1 to prove the following lemma:

LEMMA. Let  $\Sigma$  be a non-trivial closed irreducible system in  $y_1, \dots, y_n$ , and let  $B$  be a form which does not hold  $\Sigma$ . Given any positive integer  $m$ , and any solution  $\bar{y}_1(x), \dots, \bar{y}_n(x)$  of  $\Sigma$ , analytic in an open region  $\mathfrak{B}$ , let  $\mathfrak{B}_m$  be the set of all points  $x_0$  in  $\mathfrak{B}$  such that for every  $\epsilon > 0$  there is a solution  $y_1(x), \dots, y_n(x)$  of  $\Sigma$ , analytic at  $x_0$ , for which  $\mathfrak{B}$  is different from zero at  $x_0$ , and

$$|y_{ij}(x_0) - \bar{y}_{ij}(x_0)| < \epsilon, \quad i = 1, \dots, n; j = 0, 1, \dots, m. \dagger$$

Then  $\mathfrak{B} - \mathfrak{B}_m$  is isolated in  $\mathfrak{B}$ .

Let  $u_1, \dots, u_q$  be a set of arbitrary unknowns for  $\Sigma$ , let  $y_1, \dots, y_p$  be the remaining unknowns in  $\Sigma, \ddagger$  and let

$$(1) \quad A_1, \dots, A_p$$

be a basic set for  $\Sigma$  with the unknowns ordered  $u_1, \dots, u_q; y_1, \dots, y_p$ .

Let

$$(2) \quad \bar{u}_1, \dots, \bar{u}_q; \bar{y}_1, \dots, \bar{y}_p$$

be a solution of  $\Sigma$ , analytic in an open region  $\mathfrak{B}$ .

Following the procedure in ADE, §73, without change, we determine the prime algebraic system  $\Omega$ . Corresponding to (2) is a solution  $\bar{u}_{ik}, \bar{y}_{jk}$  of  $\Omega$ , analytic in  $\mathfrak{B}$ . In accordance with Theorem 1 of this paper there is an open region  $\mathfrak{M}$  in  $\mathfrak{B}$ , whose complement in  $\mathfrak{B}$  is isolated in  $\mathfrak{B}$ , such that, in every open region which with its boundary is included in a bounded simply-con-

\* Montel, *Les Familles Normales de Fonctions Analytiques*, p. 30.

† The second subscript is an index of differentiation.

‡ We renumber the unknowns if necessary.

nected subregion  $\mathfrak{C}$  of  $\mathfrak{M}$ , the solution  $\bar{u}_{ik}, \bar{y}_{jk}$  of  $\Omega$  can be approximated uniformly by solutions of  $\Omega$  for which  $BS_1 \cdots S_p$  is distinct from zero throughout  $\mathfrak{C}$ . In particular, if  $x_0$  is a point of  $\mathfrak{M}$ , then for every  $\epsilon > 0$  there is a solution  $u_{ik}, y_{jk}$  of  $\Omega$ , analytic at  $x_0$ , for which  $BS_1 \cdots S_p$  is different from zero at  $x_0$  and

$$(3) \quad \begin{aligned} |u_{ik}(x_0) - \bar{u}_{ik}(x_0)| < \epsilon, \quad |y_{jk}(x_0) - \bar{y}_{jk}(x_0)| < \epsilon, \\ i = 1, \dots, q; j = 1, \dots, p; k = 0, 1, \dots, m. \end{aligned}$$

Now the  $u_{ik}(x_0), y_{jk}(x_0)$  in (3) furnish initial conditions for a normal solution of the set of differential forms (1). Hence for every point  $x_0$  in  $\mathfrak{B}$  and every  $\epsilon > 0$  there is a solution  $u_i, y_j$  of  $\Sigma$ , analytic at  $x_0$ , which satisfies (3), and which gives  $BS_1 \cdots S_p$  a nonzero value at  $x_0$ . Thus  $\mathfrak{M}$  is included in  $\mathfrak{B}_m$ , and therefore  $\mathfrak{B} - \mathfrak{B}_m$  is isolated in  $\mathfrak{B}$ .

3. We use this lemma to prove the following theorem:

**THEOREM 2.** *Let  $\Sigma$  be a non-trivial closed irreducible system in  $y_1, \dots, y_n$ , and let  $B$  be a form which does not hold  $\Sigma$ . Then the open region\*  $\mathfrak{A}$  contains a subset  $\mathfrak{P}$  whose complement in  $\mathfrak{A}$  is at most denumerably infinite, such that for every point  $x_0$  in  $\mathfrak{P}$ , every solution  $\bar{y}_1, \dots, \bar{y}_n$  of  $\Sigma$ , analytic at  $x_0$ , every positive integer  $m$ , and every positive number  $\epsilon$  there is a solution  $y_1, \dots, y_n$  of  $\Sigma$ , analytic at  $x_0$ , for which  $B$  is different from zero at  $x_0$  and*

$$(4) \quad |y_{ij}(x_0) - \bar{y}_{ij}(x_0)| < \epsilon, \quad i = 1, 2, \dots, n; j = 0, 1, \dots, m.$$

We shall use the following notation: If  $\bar{y}_1, \dots, \bar{y}_n$  is a solution of  $\Sigma$ , analytic in an open region  $\mathfrak{B}$ , then by  $\mathfrak{P}(\bar{y}_1, \dots, \bar{y}_n; \mathfrak{B})$  we shall mean the set of points  $x_0$  in  $\mathfrak{B}$  such that for every positive integer  $m$  and every  $\epsilon > 0$  there is a solution  $y_1, \dots, y_n$  of  $\Sigma$ , analytic at  $x_0$ , for which  $B$  is different from zero at  $x_0$  and (4) holds. Now for every choice of  $\bar{y}_1, \dots, \bar{y}_n$  and  $\mathfrak{B}$  the set  $\mathfrak{P}(\bar{y}_1, \dots, \bar{y}_n; \mathfrak{B})$  has a complement in  $\mathfrak{B}$  which is at most denumerably infinite. For let  $\bar{y}_1, \dots, \bar{y}_n$  be a solution of  $\Sigma$ , analytic in  $\mathfrak{B}$ . Then for every positive integer  $m$  let  $\mathfrak{B}_m$  be the set of points  $x_0$  in  $\mathfrak{B}$  such that for every  $\epsilon > 0$  there is a solution  $y_1, \dots, y_n$  of  $\Sigma$ , analytic at  $x_0$ , for which  $\mathfrak{B}$  is different from zero at  $x_0$  and (4) holds. It is easy to see that  $\mathfrak{P}(\bar{y}_1, \dots, \bar{y}_n; \mathfrak{B})$  is identical with the intersection of the  $\mathfrak{B}_i$ , ( $i = 1, 2, \dots$ ). By the lemma just proved,  $\mathfrak{B} - \mathfrak{B}_i$ , ( $i = 1, 2, \dots$ ), is at most denumerably infinite. Consequently  $\mathfrak{B} - \mathfrak{P}(\bar{y}_1, \dots, \bar{y}_n; \mathfrak{B})$  is at most denumerably infinite.†

\* We recall that  $\mathfrak{A}$  is the open region in which are defined the functions belonging to  $\mathfrak{J}$ , the underlying field of coefficients.

† This statement is an extension of a theorem of Ritt, according to which, for every solution  $\bar{y}_1, \dots, \bar{y}_n$  of  $\Sigma$ , analytic in  $\mathfrak{B}$ , the set  $\mathfrak{P}(\bar{y}_1, \dots, \bar{y}_n; \mathfrak{B})$  is dense in  $\mathfrak{B}$ . (ADE, §74, and Ritt, *On the singular solutions of algebraic differential equations*, Annals of Mathematics, vol. 37 (1936), note 18.)

Let  $\mathfrak{R}_1, \mathfrak{R}_2, \dots$  be the set of those circles contained in  $\mathfrak{A}$  whose centers have rational coordinates and whose radii are rational. Let  $\mathfrak{S}_i$ , ( $i=1, 2, \dots$ ), be the set of all solutions of  $\Sigma$  which are analytic in the closed envelope of  $\mathfrak{R}_i$ . We consider  $\mathfrak{S}_i$  to be a metric space, the distance between two solutions  $y_1, \dots, y_n$  and  $z_1, \dots, z_n$  in  $\mathfrak{S}_i$  being given by

$$(5) \quad \delta_i((y_1, \dots, y_n), (z_1, \dots, z_n)) \\ = \max(|y_1 - z_1| + |y_2 - z_2| + \dots + |y_n - z_n|),$$

where the maximum is taken over the closed envelope of  $\mathfrak{R}_i$ . Then  $\mathfrak{S}_i$  is a separable space.\* For  $\mathfrak{S}_i$  is a subset of the separable space  $\mathcal{A}_i$  consisting of all ordered sets of  $n$  functions  $y_1, \dots, y_n$  analytic in the closed envelope of  $\mathfrak{R}_i$ , the distance between two elements  $y_1, \dots, y_n$  and  $z_1, \dots, z_n$  of  $\mathcal{A}_i$  being given by (5);  $\mathcal{A}_i$  is separable because the subset of  $\mathcal{A}_i$  consisting of all ordered sets of  $n$  polynomials with rational complex numbers for coefficients is dense in  $\mathcal{A}_i$  and denumerable.

Let  $\mathfrak{C}_i$  be a denumerable dense subset of  $\mathfrak{S}_i$ , ( $i=1, 2, \dots$ ). Let  $\Omega$  be the set-theoretic sum

$$\sum (\mathfrak{R}_i - \mathfrak{P}(\bar{y}_1, \dots, \bar{y}_n; \mathfrak{R}_i)),$$

where  $i$  ranges over the positive integers, and for each  $i$  the solution  $\bar{y}_1, \dots, \bar{y}_n$  ranges over  $\mathfrak{C}_i$ . Then  $\Omega$  is at most denumerably infinite. We define  $\mathfrak{P}$  as the complement of  $\Omega$  in  $\mathfrak{A}$ . Let  $x_0$  be a point of  $\mathfrak{P}$ ,  $\bar{y}_1, \dots, \bar{y}_n$  a solution of  $\Sigma$ , analytic at  $x_0$ ,  $m$  a positive integer, and  $\epsilon$  a positive number. Then there is a  $\mathfrak{R}_i$  containing  $x_0$  such that  $\bar{y}_1, \dots, \bar{y}_n$  is analytic in the closed envelope of  $\mathfrak{R}_i$ . There exists a solution  $\bar{y}_1, \dots, \bar{y}_n$  belonging to  $\mathfrak{C}_i$  such that

$$(6) \quad |\bar{y}_{ij}(x_0) - \bar{y}_{ij}(x_0)| < \epsilon/2, \quad i = 1, \dots, n; j = 0, 1, \dots, m,$$

since there is a sequence of solutions in  $\mathfrak{C}_i$  convergent to  $\bar{y}_1, \dots, \bar{y}_n$  uniformly in  $\mathfrak{R}_i$ . Since  $x_0$  is in  $\mathfrak{P}(\bar{y}_1, \dots, \bar{y}_n; \mathfrak{R}_i)$ , there is a solution  $y_1, \dots, y_n$  of  $\Sigma$ , analytic at  $x_0$ , for which  $\mathfrak{P}$  is different from zero at  $x_0$  and

$$(7) \quad |y_{ij}(x_0) - \bar{y}_{ij}(x_0)| < \epsilon/2, \quad i = 1, \dots, n; j = 0, 1, \dots, m.$$

By (6) and (7) we have (4). Since  $\mathfrak{A} - \mathfrak{P}$  is at most denumerably infinite, we have our theorem.

## PART II. SEQUENCES OF IRREDUCIBLE SYSTEMS

### 4. We state first the following lemma:

LEMMA 1. Let  $\Sigma$  be a prime system in the unknowns  $u_1, \dots, u_q; y_1, \dots, y_p$ , where  $u_1, \dots, u_q$  is a set of unconditioned unknowns for  $\Sigma$ , and let

\* Of course  $\mathfrak{S}_i$  may be an empty or finite set.

$$(8) \quad A_1, \dots, A_p$$

be a basic set for  $\Sigma$ , with  $A_i$  introducing  $y_i$ , ( $i=1, \dots, p$ ). Let  $F$  be a simple form which does not hold  $\Sigma$ . Then there exists a nonzero simple form  $G$  in  $u_1, \dots, u_q$  which is a linear combination of the simple forms  $A_1, \dots, A_p, F$ .

Let  $J$  be the product of the initials in (8). Let  $K$  be a nonzero simple form in  $u_1, \dots, u_q$  such that every solution of (8) which annuls  $J$  is a solution of  $K$ .\* Let  $\Phi$  be the system of simple forms  $A_1, \dots, A_p, F$ , and let  $\Phi$  be equivalent to the prime systems  $\Pi_1, \dots, \Pi_r; \Lambda_1, \dots, \Lambda_s$  where every  $\Pi_i$  is held by  $K$  and no  $\Lambda_i$  is held by  $K$ . It is easy to see that each  $\Lambda_i$  is held by  $\Sigma + F$  and is therefore of lower dimensionality than  $\Sigma$ ; that is, has a nonzero simple form  $G_i$  in  $u_1, \dots, u_q$ .

Set  $H = G_1 \dots G_r K$ . Then  $H$  is a nonzero simple form in  $u_1, \dots, u_q$  which holds  $\Phi$ . Consequently, there is a positive integer  $\sigma$  such that  $H^\sigma$  is a linear combination of the simple forms of  $\Phi$ . We evidently may take  $H^\sigma$  for the simple form  $G$  whose existence is to be demonstrated.

5. We can now prove the following lemma:

LEMMA 2. Let  $\Lambda$  be a prime system in the unknowns  $v_1, \dots, v_t; z_1, \dots, z_s$ , where  $v_1, \dots, v_t$  is a set of unconditioned unknowns for  $\Lambda$ , and let  $C_1, \dots, C_s$  be a basic set for  $\Lambda$ , with  $C_i$  introducing  $z_i$ , ( $i=1, \dots, s$ ). Let  $T_i$  be the separant of  $C_i$ , and let  $F$  be any form which does not hold  $\Lambda$ . Then there is a set  $\mathfrak{R}$  in  $\mathfrak{A}$  with the following properties:

(i)  $\mathfrak{A} - \mathfrak{R}$  is isolated in  $\mathfrak{A}$ .

(ii) If  $\sigma$  is any integer with  $1 \leq \sigma \leq s$ , if  $x_0$  is any point of  $\mathfrak{R}$ , and if  $a_1, \dots, a_t; b_1, \dots, b_s$  is a set of complex numbers such that  $C_i = 0$  and  $T_i \neq 0$ , ( $i=1, \dots, \sigma$ ), when  $x = x_0$ ,  $v_j = a_j$ , and  $z_k = b_k$ , ( $j=1, \dots, t; k=1, \dots, \sigma$ ), then for every  $\delta > 0$  there is a solution  $v_1(x), \dots, v_t(x); z_1(x), \dots, z_s(x)$  of  $\Lambda$ , analytic at  $x_0$ , satisfying the inequalities

$$|v_j(x_0) - a_j| < \delta, \quad |z_k(x_0) - b_k| < \delta, \quad j = 1, \dots, t; k = 1, \dots, \sigma,$$

and giving  $F$  a nonzero value at  $x_0$ .†

Let  $I_i$  be the initial of  $C_i$ , ( $i=1, \dots, s$ ), and let  $k$  be any integer with  $1 \leq k \leq s-1$ . Now  $C_1, \dots, C_k$  is a basic set for a prime system  $\Lambda_k$  which has  $v_1, \dots, v_t$  for a set of unconditioned unknowns.‡  $I_{k+1}$  does not hold  $\Lambda_k$ , since it does not hold  $\Lambda$ . By Lemma 1 there is an identity

$$(9) \quad G_k = E_{k,1}C_1 + E_{k,2}C_2 + \dots + E_{k,k}C_k + E_{k,k+1}I_{k+1},$$

\* Ritt, *Systems of algebraic differential equations*, Annals of Mathematics, (2), vol. 36 (1935), §8.

† We shall apply this lemma only for the case  $\sigma=1$ .

‡ Cf. ADE, §45.

where  $G_k$  is a nonzero simple form in  $v_1, \dots, v_t$ . Likewise, since  $T_1 \dots T_s F$  does not hold  $\Lambda$ , there is an identity

$$(10) \quad G = F_1 C_1 + F_2 C_2 + \dots + F_s C_s + F_{s+1} T_1 \dots T_s F,$$

where  $G$  is a nonzero simple form in  $v_1, \dots, v_t$ . Set  $L = I_1 G G_1 \dots G_{s-1}$ . Then  $L$  is a nonzero simple form in  $v_1, \dots, v_t$ . We now present the set  $\mathfrak{N}$ . Let  $\mathfrak{N}$  consist of all points in  $\mathfrak{A}$  at which the coefficients of the simple forms appearing in the identities (10) and (9), ( $k=1, \dots, s-1$ ), are analytic, and at which  $L$  has one or several nonzero coefficients. Evidently  $\mathfrak{A} - \mathfrak{N}$  is isolated in  $\mathfrak{A}$ . Now let  $\sigma$  be any positive integer with  $1 \leq \sigma \leq s$ , let  $x_0$  be any point of  $\mathfrak{N}$ , assume  $\delta > 0$ , and let  $a_1, \dots, a_t; b_1, \dots, b_s$  be complex numbers such that  $C_i = 0$  and  $T_i \neq 0$ , ( $i=1, \dots, \sigma$ ), when  $x=x_0, v_j=a_j, z_k=b_k$ , ( $j=1, \dots, t; k=1, \dots, \sigma$ ). Since  $T_1 \dots T_\sigma$  is equal to the Jacobian  $\partial(C_1, \dots, C_\sigma)/\partial(z_1, \dots, z_\sigma)$ , there is a unique set of functions  $f_1(x, v_1, \dots, v_t), \dots, f_\sigma(x, v_1, \dots, v_t)$ , analytic near  $(x_0, a_1, \dots, a_t)$  such that  $b_i = f_i(x_0, a_1, \dots, a_t)$ , ( $i=1, \dots, \sigma$ ), and such that the substitution of  $f_i(x, v_1, \dots, v_t)$  for  $z_i$  in  $C_1, \dots, C_\sigma$  yields  $\sigma$  functions of  $x, v_1, \dots, v_t$  each of which is identically zero.

Let  $\mathfrak{N}$  be a neighborhood of  $(x_0, a_1, \dots, a_t)$  in which every  $f_i(x, v_1, \dots, v_t)$  is analytic ( $i=1, \dots, \sigma$ ) such that for every point  $(x_1, c_1, \dots, c_t)$  in  $\mathfrak{N}$  the relations

$$|c_j - a_j| < \delta, \quad |f_i(x_0, c_1, \dots, c_t) - b_i| < \delta, \quad i=1, \dots, \sigma,$$

are valid. Let  $c_1, \dots, c_t$  be chosen so that  $(x_0, c_1, \dots, c_t)$  is a point of  $\mathfrak{N}$  at which  $L$  is not zero. Such a point exists because  $L$  has one or several coefficients different from zero at  $x_0$ . Let  $d_i = f_i(x_0, c_1, \dots, c_t)$ , ( $i=1, \dots, \sigma$ ). Then the substitution

$$(11) \quad x = x_0, v_j = c_j, z_i = d_i, \quad i=1, \dots, \sigma; j=1, \dots, t,$$

annuls  $C_1, \dots, C_\sigma$  but not  $G_\sigma$ , and therefore does not annul  $I_{\sigma+1}$  (by (9), with  $k=\sigma$ ). Therefore the polynomial in  $z_{\sigma+1}$  obtained from  $C_{\sigma+1}$  by the substitution (11) has at least one root  $d_{\sigma+1}$ . The substitution

$$(12) \quad x = x_0, v_j = c_j, z_i = d_i, \quad i=1, \dots, \sigma+1; j=1, \dots, t,$$

annuls  $C_1, \dots, C_{\sigma+1}$  but not  $G_{\sigma+1}$ , and therefore does not annul  $I_{\sigma+2}$  (by (9), with  $k=\sigma+1$ ). Hence the polynomial in  $z_{\sigma+2}$  obtained from  $C_{\sigma+2}$  by the substitution (12) has at least one root  $d_{\sigma+2}$ .

Continuing in this manner, we obtain a set of values  $c_1, \dots, c_t; d_1, \dots, d_s$  such that the substitution

$$(13) \quad x = x_0, v_j = c_j, z_i = d_i, \quad i=1, \dots, s; j=1, \dots, t,$$

annuls  $C_1, \dots, C_s$  but not  $G$ , and therefore does not annul  $T_1 \dots T_s F$  (by

(10)). Since  $T_1 \cdots T_s = \partial(C_1, \cdots, C_s) / \partial(z_1, \cdots, z_s)$ , there is a unique set of functions  $\zeta_1(x, v_1, \cdots, v_t), \cdots, \zeta_s(x, v_1, \cdots, v_t)$ , analytic near  $(x_0, c_1, \cdots, c_t)$ , such that  $d_i = \zeta_i(x_0, c_1, \cdots, c_t)$ , ( $i = 1, \cdots, s$ ), and such that the substitution  $z_i = \zeta_i(x, v_1, \cdots, v_t)$ , ( $i = 1, \cdots, s$ ), transforms  $C_1, \cdots, C_s$  into  $s$  functions of  $x, v_1, \cdots, v_t$ , each of which is identically zero.

Let  $v_j(x) = c_j$ ,  $z_i(x) = \zeta_i(x, c_1, \cdots, c_t)$ , ( $j = 1, \cdots, t$ ;  $i = 1, \cdots, s$ ). Then  $v_j(x), z_i(x)$  is evidently a solution of  $\Lambda$ , analytic at  $x_0$ , for which  $F$  is different from zero at  $x_0$ , and

$$|v_j(x_0) - a_j| < \delta, \quad |z_k(x_0) - b_k| < \delta, \quad j = 1, \cdots, t; \quad k = 1, \cdots, \sigma.$$

6. We now prove the following theorem:

THEOREM 3. *Let*

$$(14) \quad \Sigma_1, \Sigma_2, \cdots$$

*be a sequence of closed irreducible systems in the unknowns  $y_1, \cdots, y_n$  such that the manifold of  $\Sigma_i$  is a proper part of the manifold of  $\Sigma_{i+1}$ , ( $i = 1, 2, \cdots$ ). Let  $\Sigma$  be the set of all forms  $F$  such that  $F$  is in every system  $\Sigma_i$ , ( $i = 1, 2, \cdots$ ). Then  $\Sigma$  is a closed irreducible system having more arbitrary unknowns than  $\Sigma_i$ , ( $i = 1, 2, \cdots$ ).*\*

$\Sigma$  is obviously closed.  $\Sigma$  is irreducible because if  $GH$  holds  $\Sigma$ , then either  $G$  holds an infinite set of the  $\Sigma_i$ , hence all the  $\Sigma_i$ , or  $H$  does; so either  $G$  is in  $\Sigma$  or  $H$  is.

Now let  $\Sigma_i$  be any system in (14). Evidently, if there is a set of unknowns in which  $\Sigma_i$  has no nonzero form, then  $\Sigma$  has no nonzero form in the unknowns of that set. Hence  $\Sigma$  has at least as many arbitrary unknowns as  $\Sigma_i$ . Now suppose that there is an  $m$  such that  $\Sigma_m$  has the same number of arbitrary unknowns as  $\Sigma$ . Then there is a set of unknowns  $y_{i_1}, y_{i_2}, \cdots, y_{i_q}$  which is a set of arbitrary unknowns for  $\Sigma$ , and which is also a set of arbitrary unknowns for  $\Sigma_m$ . Now  $\Sigma_j$ , ( $j \geq m$ ), has no nonzero form in  $y_{i_1}, \cdots, y_{i_q}$  because  $\Sigma_m$  has no such form. But  $\Sigma_j$  has not more than  $q$  arbitrary unknowns, because  $\Sigma$  has  $q$  arbitrary unknowns. Hence  $y_{i_1}, \cdots, y_{i_q}$  is a set of arbitrary unknowns for every  $\Sigma_j$ , ( $j \geq m$ ). Taking  $y_{i_1}, \cdots, y_{i_q}$  as a set of arbitrary unknowns for  $\Sigma$  and for each  $\Sigma_j$ , ( $j \geq m$ ), we introduce a resolvent for  $\Sigma$  and for each  $\Sigma_j$  (adjoining  $x$  to  $\mathcal{F}$  if necessary) and we let  $\rho, \rho_j$  be the orders of the resolvents of  $\Sigma, \Sigma_j$ , respectively ( $j \geq m$ ). By a theorem of E. Gourin,† since  $\Sigma$  has the same

\* For a closed irreducible system  $\Lambda$ , the number of unknowns in a set of arbitrary unknowns for  $\Lambda$  is independent of the manner in which the set is chosen (ADE, §30). This number we call the *number of arbitrary unknowns in  $\Lambda$* .

† E. Gourin, *On irreducible systems of algebraic differential equations*, Bulletin of the American Mathematical Society, vol. 39 (1933), pp. 593-595.

set of arbitrary unknowns as each of the  $\Sigma_j$ , we have  $\rho_m < \rho_{m+1} < \dots < \rho$ , which is clearly impossible.

COROLLARY.  $\Sigma$  has a non-empty set of arbitrary unknowns.

7. We now make the following definition:

DEFINITION. Let  $n$  be any positive integer. Let  $f_1(x), \dots, f_n(x)$  be an ordered set of  $n$  functions analytic in an open region  $\mathfrak{B}$ . Let  $\mathcal{N}$  be a set each of whose elements is an ordered set of  $n$  functions which have a region of analyticity in common. Then if  $m$  is a positive integer, we shall say that a point  $x_0$  of  $\mathfrak{B}$  is a point of  $m$ th order contact between the set  $f_1(x), \dots, f_n(x)$  and the set of sets  $\mathcal{N}$  if for every  $\epsilon > 0$  there is a set  $y_1(x), \dots, y_n(x)$  in  $\mathcal{N}$ , analytic at  $x_0$ , such that

$$|y_{ij}(x_0) - f_{ij}(x_0)| < \epsilon, \quad i = 1, \dots, n; j = 0, 1, \dots, m.$$

If  $x_0$  is a point of  $m$ th order contact between  $f_1(x), \dots, f_n(x)$  and  $\mathcal{N}$  for every  $m$ , then we shall say simply that  $x_0$  is a point of contact between  $f_1(x), \dots, f_n(x)$  and  $\mathcal{N}$ .

8. We can now state the following theorem:

THEOREM 4. Let  $\Sigma_1, \Sigma_2, \dots$ , and  $\Sigma$  be as in the hypothesis of Theorem 3. Let  $\mathcal{M}_i$  be the manifold of  $\Sigma_i$ , ( $i=1, 2, \dots$ ), and let  $\mathcal{N}$  be the set-theoretic sum  $\mathcal{M}_1 + \mathcal{M}_2 + \dots$ . If  $f_1(x), \dots, f_n(x)$  is an ordered set of  $n$  functions analytic in an open region  $\mathfrak{B}$ , then a necessary and sufficient condition for  $f_1(x), \dots, f_n(x)$  to be a solution of  $\Sigma$  is that  $\mathfrak{B}$  contain a point of contact between  $f_1, \dots, f_n$  and  $\mathcal{N}$ .

Sufficiency proof. Let

$$(15) \quad f_1(x), \dots, f_n(x)$$

be an ordered set of  $n$  functions analytic in an open region  $\mathfrak{B}$  which contains a point  $x_0$  of contact between (15) and  $\mathcal{N}$ . We shall prove that (15) is a solution of  $\Sigma$ .

If  $H$  is a form in  $\Sigma$  whose coefficients are analytic at  $x_0$ , then  $H$ , considered as a function of  $x$  and the letters appearing in  $H$ , is continuous when  $x$  is near  $x_0$ .

Let  $m$  be a positive integer greater than the order of  $H$  in  $y_i$ , ( $i=1, \dots, n$ ). Assume  $\epsilon > 0$ . Let

$$(16) \quad y_1(x), \dots, y_n(x)$$

be a solution in  $\mathcal{N}$ , analytic at  $x_0$ , such that

$$(17) \quad |f_{ij}(x_0) - y_{ij}(x_0)| < \epsilon, \quad i = 1, \dots, n; j = 0, 1, \dots, m.$$

\* The second subscript is an index of differentiation.

When we substitute (15) and (16) in  $H$ , we obtain functions  $h(x)$  and  $k(x)$ , respectively, which are analytic near  $x_0$ . Evidently  $k(x) \equiv 0$ , since (16) is in  $\mathcal{N}$  and is therefore a solution of  $\Sigma$ . In particular  $k(x_0) = 0$ . Now  $\epsilon$  is arbitrarily small; consequently, the relations (17) and the continuity of  $H$  imply that  $|k(x_0) - h(x_0)|$  is arbitrarily small. Thus  $|h(x_0)|$  is arbitrarily small, so that  $h(x_0) = 0$ . Now  $h'(x_0)$  may be obtained by substituting (15) in  $H'$  and putting  $x = x_0$ .<sup>\*</sup> Hence, by the preceding argument,  $h'(x_0) = 0$ . Continuing in this manner we prove that every derivative of  $h(x)$  vanishes at  $x_0$ . This means  $h(x) \equiv 0$ . Hence (15) is a solution of  $H$ , where  $H$  is any form of  $\Sigma$  whose coefficients are analytic at  $x_0$ . But if  $G$  is any form of  $\Sigma$ , the product of  $G$  by a suitable nonzero function  $\psi(x)$  in  $\mathcal{F}$  is a form  $H$  of  $\Sigma$  with coefficients analytic at  $x_0$ .<sup>†</sup>

Since (15) annuls  $H$ , it annuls  $G$ . This proves that (15) is a solution of  $\Sigma$ .<sup>‡</sup>

**Necessity proof.** The necessity of the condition is implied by Theorem 6, below.

9. The next theorem is as follows:

**THEOREM 5.** *Let the notations  $\Sigma_i$ ,  $\mathcal{M}_i$ , and  $\Sigma$  have the same significance as in the hypothesis of Theorem 4. Then there exists a function  $b = b(m)$ , defined on the positive integers, and assuming positive integral values, such that for every solution  $y_1(x), \dots, y_n(x)$  of  $\Sigma$ , analytic in an open region  $\mathfrak{B}$ , and every positive integer  $m$ , the set of points of  $m$ th order contact between  $y_1(x), \dots, y_n(x)$  and  $\mathcal{M}_{b(m)}$  is a set whose complement in  $\mathfrak{B}$  is isolated in  $\mathfrak{B}$ .*

Let  $u_1, \dots, u_q$  be a set of arbitrary unknowns for  $\Sigma$ . If  $\Sigma$  is non-trivial, we introduce a resolvent for  $\Sigma$  with a new unknown  $\omega$  satisfying  $\omega - Q = 0$ .<sup>§</sup> If  $\Sigma$  is trivial (that is, if  $\Sigma$  has no nonzero forms) we introduce a new unknown  $\omega$  satisfying  $\omega - Q = 0$ , where  $Q \equiv 0$ .

In either case let  $\Omega$  be the set of forms holding  $\Sigma + (\omega - Q)$ , and let  $\Omega_i$  be the set of forms holding  $\Sigma_i + (\omega - Q)$ , ( $i = 1, 2, \dots$ ). Then  $\Omega_{i+1}$  holds  $\Omega_i$ ,  $\Omega$  is closed and irreducible,<sup>||</sup> and  $\Omega$  is the set of all forms  $F$  such that  $F$  is in every system  $\Omega_i$ , ( $i = 1, 2, \dots$ ).

Let  $y_1, \dots, y_p$  be the unknowns in  $\Omega$  other than  $u_1, \dots, u_q$ ;  $\omega$ .<sup>¶</sup> Let the unknowns be ordered  $u_1, \dots, u_q; \omega; y_1, \dots, y_p$ , and let

$$(18) \quad R, A_1, \dots, A_p$$

<sup>\*</sup> Superscripts indicate differentiation.

<sup>†</sup> For if  $G$  has a coefficient  $\phi(x)$  with a pole at  $x_0$ , then the reciprocal of  $\phi(x)$  will have a zero at  $x_0$ , and a suitable power of that reciprocal will serve as  $\psi(x)$ .

<sup>‡</sup> This proof is similar to a proof given by Ritt for a different theorem, ADE, §72.

<sup>§</sup> ADE, §§25-29.

<sup>||</sup>  $\Omega_i$  is also closed and irreducible.

<sup>¶</sup> We renumber the unknowns if necessary.

be a corresponding basic set for  $\Omega$ . Then  $A_i$  is of zero order in  $y_i$  and is linear in  $y_m$ , ( $i=1, \dots, p$ ).\*

Let  $h$  be the order of  $R$  in  $\omega$ . We assert that if  $a$  is any positive integer, then there is an integer  $b$  depending upon  $a$  such that the system  $\Omega_b$  has no nonzero form in the letters

$$(19) \quad u_{\alpha\beta}, \omega_\gamma, \quad \alpha = 1, \dots, q; \beta = 0, 1, \dots, a; \gamma = 0, 1, \dots, h-1.$$

For let us assume that this assertion is false. Then there is an  $a$  such that every  $\Omega_i$  has a nonzero form in the letters (19). From each  $\Omega_i$  let a nonzero form  $F_i$  in the letters (19) be selected which is of minimum rank. Without loss of generality we may assume that  $F_i$  is algebraically irreducible ( $i=1, 2, \dots$ ). Since  $\Omega_2$  must have solutions, each  $F_i$ , ( $i>2$ ), involves unknowns.

Since the totality of letters involved in the  $F_i$  is a finite set, there is an infinite subset of the  $F_i$  such that if  $F_k$  and  $F_l$  are two forms in the subset, then  $F_k$  and  $F_l$  have the same order in  $u_\alpha$ , ( $\alpha=1, \dots, q$ ), and the same order in  $\omega$ . We assert that the quotient of any two forms in this subset is a (nonzero) function in  $\mathcal{F}$ . For if  $F_k$  and  $F_l$ , ( $k < l$ ), are relatively prime, then the resultant  $G$  of  $F_k$  and  $F_l$ , with respect to the highest letter in  $F_k$  and  $F_l$ , is a nonzero form free of that letter. Then  $\Omega_k$  has the form  $G$  in the letters (19).† But  $G$  is lower than  $F_k$ . This contradiction with the minimal property of  $F_k$  proves that there is an infinite set of the  $F_i$ , each of which is the product of a fixed  $F_k$  by a nonzero function in  $\mathcal{F}$ . Then this  $F_k$  is in all the  $\Omega_i$ , and therefore in  $\Omega$ , although it is lower than  $R$ . This contradiction proves that for every  $a$  there is a  $b$  such that  $\Omega_b$  has no nonzero form in the letters (19).

Now let  $S, S_i$  be the separant of  $R, A_i$ , ( $i=1, \dots, p$ ), respectively, and define  $K_1 \equiv SS_1 \cdots S_p$ . Discarding a finite set of the  $\Omega_i$  if necessary, we assume that  $K_1$  holds no  $\Omega_i$ , since  $K_1$  is not in  $\Omega$ .

Let  $g$  be a positive integer greater than the maximum order of each form of (18) in each unknown.

Let  $m$  be any positive integer, to be fixed throughout the remainder of this proof. Let  $a = m + g$ , and let  $b = b(m)$  be the smallest positive integer such that  $\Omega_b$  has no nonzero form in the letters (19).

We take any set of arbitrary unknowns for  $\Omega_b$ , order the remaining unknowns in any fashion, and let

$$(20) \quad B_1, B_2, \dots, B_r$$

be a corresponding basic set for  $\Omega_b$ .

\* Since either  $R$  is a resolvent for  $\Sigma$ , or  $\Sigma$  is trivial. In the latter case (18) is simply the form  $R$ .

† Since  $G$  is a linear combination of  $F_k$  and  $F_l$ , each of which is in  $\Omega_k$ .

Let  $\tau, \tau_i, \sigma_j$  be the orders of the highest derivatives of  $\omega, y_i, u_j$  ( $i = 1, \dots, p$ ;  $j = 1, \dots, q$ ), respectively, appearing in (20).<sup>\*</sup> Let  $K_2$  be the product of the separants in (20). Let  $\Lambda$  be the set of all simple forms that vanish for all solutions of the system

$$(21) \quad B_{ij}, \quad i = 1, \dots, r; j = 0, 1, \dots, a, \dagger$$

for which  $K_2 \neq 0$ , where the forms (21) are to be considered as simple forms in the unknowns

$$(22) \quad \begin{array}{ll} u_{ij}, \omega_k, y_{\mu\nu}, & j = 0, 1, \dots, a + \sigma_i; i = 1, \dots, q; \\ k = 0, 1, \dots, a + \tau; & \nu = 0, 1, \dots, a + \tau_\mu; \mu = 1, \dots, p. \end{array}$$

It is easy to see that  $\Lambda$  is prime, that every simple form which holds  $\Lambda$ , when considered as a form in the unknowns  $u_1, \dots, u_q; \omega; y_1, \dots, y_p$ , and their derivatives, will hold  $\Omega_b$ , and that every form of  $\Omega_b$  in the letters (22), when considered as a simple form in those letters, will hold  $\Lambda$ .<sup>‡</sup>

Since every form in  $\Lambda$  is in  $\Omega_b$ , there is no nonzero form in  $\Lambda$  in the letters (19). Renaming the letter (22), let

$$(23) \quad v_1, \dots, v_t$$

be a set of unconditioned unknowns for  $\Lambda$ , and let

$$(24) \quad z_1, \dots, z_s$$

be the other unknowns in  $\Lambda$ . We may and do choose (23), (24) so that (23) includes (19) and also so that  $z_1 = \omega_h$ , the latter being possible because  $R$  is in  $\Lambda$ .

With the unknowns ordered  $v_1, \dots, v_t; z_1, \dots, z_s$ , let

$$(25) \quad C_1, C_2, \dots, C_s$$

be a basic set for  $\Lambda$ . Then  $R$  can be taken for  $C_1$ . For if  $F$  were a simple form of  $\Lambda$  in the unknowns  $v_1, \dots, v_t; \omega_h$ , of lower degree than  $R$  in  $\omega_h$ , then the resultant of  $R$  and  $F$  with respect to  $\omega_h$  would be a nonzero simple form of  $\Lambda$  in the letters (23),<sup>§</sup> although (23) is a set of unconditioned unknowns. We shall assume that  $C_1 = R$ .

<sup>\*</sup> If  $\omega, y_i$ , or  $u_j$  does not appear in (20), then  $\tau, \tau_i$ , or  $\sigma_j$ , respectively, is to be taken as zero.

<sup>†</sup> The second subscript is an index of differentiation.

<sup>‡</sup> If  $F$  holds  $\Lambda$ , then  $K_2 F$  vanishes for every solution of  $\Omega_b$ , since such a solution either annuls  $K_2$  or yields a solution of (21) for which  $K_2 \neq 0$ ; hence  $F$  holds  $\Omega_b$ , since  $K_2$  does not. Conversely, if  $F$  holds  $\Omega_b$ , then  $F$  vanishes for every solution of (21) for which  $K_2 \neq 0$ , since such a solution provides, at every point where the coefficients in (20) are analytic and  $K_2 \neq 0$ , initial conditions for a normal solution of (20). If  $FG$  holds  $\Lambda$ , then it holds  $\Omega_b$ ; so either  $F$  holds  $\Omega_b$ , hence  $\Lambda$ , or  $G$  does. Thus  $\Lambda$  is indecomposable.  $\Lambda$  is obviously simply closed.

<sup>§</sup>  $R$  is algebraically irreducible as a polynomial in  $\omega_h$ , in the field  $\mathcal{F}(v_1, \dots, v_t)$ . Cf. ADE, §§65, 45.

We note that  $K_2$  does not hold  $\Lambda$ .

Taking  $K_2$  for the form  $F$  in the hypothesis of Lemma 2, we let  $\mathfrak{N}$  be the corresponding point set in  $\mathfrak{A}$  with the properties (i), (ii) of the lemma.

Let  $\mathfrak{N}_m$  be the set of points in  $\mathfrak{N}$  at which the coefficients of the forms in (18) and (20) are analytic. Evidently  $\mathfrak{A} - \mathfrak{N}_m$  is isolated in  $\mathfrak{A}$ .

Let  $x_0$  be a point of  $\mathfrak{N}_m$ , and let

$$(26) \quad \bar{u}_1(x), \dots, \bar{u}_q(x); \bar{\omega}(x); \bar{y}_1(x), \dots, \bar{y}_p(x)$$

be a normal solution of (18), analytic at  $x_0$  and giving  $K_1$  a nonzero value at  $x_0$ . We shall prove that  $x_0$  is a point of  $m$ th order contact between (26) and the manifold of  $\Omega_b$ . Assume

$$(27) \quad R_k \equiv S\omega_{h+k} + V_k, \quad A_{ij} \equiv S_i y_{ij} + T_{ij}, \\ k = 1, 2, \dots, m; i = 1, 2, \dots, p; j = 0, 1, \dots, m.*$$

Then  $V_k$  is of order less than  $h+k$  in  $\omega$ , and  $T_{ij}$  is of order less than  $j$  in  $y_i$ . Evidently the equations

$$(28) \quad R_k = 0, A_{ij} = 0, \quad k = 1, 2, \dots, m; i = 1, \dots, p; j = 0, 1, \dots, m,$$

define  $\omega_{h+k}$ ,  $y_{ij}$  recursively as functions of  $x$ ,  $\omega_h$ , and the letters (19), continuous near  $(x_0, \bar{u}_{\alpha\beta}(x_0), \bar{\omega}_\gamma(x_0), \bar{\omega}_h(x_0))$ , ( $\alpha=1, \dots, q; \beta=0, 1, \dots, a; \gamma=0, 1, \dots, h-1$ ).†

Assume  $\epsilon > 0$ . Then there is a  $\delta > 0$  such that if  $\bar{u}_1(x), \dots, \bar{u}_q(x); \bar{\omega}(x); \bar{y}_1(x), \dots, \bar{y}_p(x)$  is a solution of  $\Omega_b$  with

$$(29) \quad |\bar{u}_{\alpha\beta}(x_0) - \bar{u}_{\alpha\beta}(x_0)| < \delta, \quad |\bar{\omega}_\gamma(x_0) - \bar{\omega}_\gamma(x_0)| < \delta, \quad |\bar{\omega}_h(x_0) - \bar{\omega}_h(x_0)| < \delta,$$

then

$$(30) \quad |\bar{u}_{jk}(x_0) - \bar{u}_{jk}(x_0)| < \epsilon, \quad |\bar{\omega}_k(x_0) - \bar{\omega}_k(x_0)| < \epsilon, \\ |\bar{y}_{ik}(x_0) - \bar{y}_{ik}(x_0)| < \epsilon, \\ j = 1, \dots, q; i = 1, \dots, p; k = 0, 1, \dots, m.$$

This results from the fact that every solution of  $\Omega$ , and therefore also every solution of  $\Omega_b$ , must satisfy the equations (28).

Now each  $v_j$ , ( $j=1, \dots, t$ ), corresponds to one of the letters (22); let  $a_j$  be the value at  $x_0$  assigned to that letter by (26). Let  $b_1 = \bar{\omega}_h(x_0)$ .

Then  $C_1$ , ( $C_1 = R$ ), vanishes under the substitution  $x = x_0$ ,  $v_j = a_j$ ,  $\omega_h = b_1$ , ( $j=1, \dots, t$ ), and  $S$  does not; thus, since  $x_0$  is in  $\mathfrak{N}$ , there is a solution  $v_j(x)$ ,

\*  $R_k$  is the  $k$ th derivative of  $R$ , and  $A_{ij}$  is the  $j$ th derivative of  $A_i$ .

† Henceforth, whenever  $\alpha, \beta, \gamma$  appear as subscripts, we shall understand that their ranges are the ones given here.

$z_i(x)$ , ( $i=1, \dots, s; j=1, \dots, t$ ), of  $\Lambda$ , analytic at  $x_0$ , for which  $|v_j(x_0) - a_j| < \delta$ ,  $|z_1(x_0) - b_1| < \delta$ , ( $j=1, \dots, t$ ), and for which  $K_2$  is different from zero at  $x_0$ .

Evidently this solution of  $\Lambda$  provides initial conditions at  $x_0$  for a normal solution  $\bar{u}_1, \dots, \bar{y}_p$  of (20) which satisfies (29). The inequalities (30) are valid for this solution of  $\Omega_b$ . Therefore  $x_0$  is a point of  $m$ th order contact between (26) and the manifold of  $\Omega_b$ .

Now let

$$(31) \quad u_1(x), \dots, u_q(x); y_1(x), \dots, y_p(x)$$

be a solution of  $\Sigma$ , analytic in an open region  $\mathfrak{B}$ . Corresponding to (31) is a solution

$$(32) \quad u_1(x), \dots, u_q(x); \omega(x); y_1(x), \dots, y_p(x)$$

of  $\Omega$ , analytic in  $\mathfrak{B}$ . According to the lemma of §2 there is a set  $\mathfrak{B}_m$  whose complement in  $\mathfrak{B}$  is isolated in  $\mathfrak{B}$ , such that every point  $x_0$  in  $\mathfrak{B}_m$  is a point of  $m$ th order contact between (32) and the set of those solutions of  $\Omega$  which give  $K_1$  a nonzero value at  $x_0$ . Let  $\mathfrak{R}_m = \mathfrak{B}_m \cdot \mathfrak{R}_m$ . Then for every  $\epsilon > 0$  and every point  $x_0$  in  $\mathfrak{R}_m$  there is a solution

$$(33) \quad \bar{u}_1(x), \dots, \bar{u}_q(x); \bar{\omega}(x); \bar{y}_1(x), \dots, \bar{y}_p(x)$$

of  $\Omega$ , analytic at  $x_0$ , for which

$$(34) \quad \begin{aligned} &|\bar{u}_{jk}(x_0) - u_{jk}(x_0)| < \epsilon, \quad |\bar{\omega}_k(x_0) - \omega_k(x_0)| < \epsilon, \\ &|\bar{y}_{ik}(x_0) - y_{ik}(x_0)| < \epsilon, \\ &i = 1, \dots, p; j = 1, \dots, q; k = 0, 1, \dots, m, \end{aligned}$$

and for which  $K_1 \neq 0$  at  $x_0$ ; and then there is a solution

$$(35) \quad \bar{u}_1(x), \dots, \bar{u}_q(x); \bar{\omega}(x); \bar{y}_1(x), \dots, \bar{y}_p(x)$$

of  $\Omega_b$  for which (30) holds. By (30) and (34) we have in particular

$$(36) \quad \begin{aligned} &|\bar{u}_{jk}(x_0) - u_{jk}(x_0)| < 2\epsilon, \quad |\bar{y}_{ik}(x_0) - y_{ik}(x_0)| < 2\epsilon, \\ &i = 1, \dots, q; j = 1, \dots, p; k = 0, 1, \dots, m. \end{aligned}$$

Consequently every point of  $\mathfrak{R}_m$  is a point of  $m$ th order contact between (31) and  $\mathfrak{M}_b$ . Since  $\mathfrak{B} - \mathfrak{R}_m$  is evidently isolated in  $\mathfrak{B}$ , we have our theorem.

10. As a consequence of Theorem 5 we have the following theorem:

**THEOREM 6.** *Let the notations  $\Sigma_i$ ,  $\mathfrak{M}_i$ ,  $\Sigma$ ,  $\mathfrak{N}$  have the same significance as in Theorem 4. Then the open region  $\mathfrak{A}$  contains a subset  $\mathfrak{B}$  whose complement in  $\mathfrak{A}$  is at most denumerably infinite, such that if  $x_0$  is a point in  $\mathfrak{B}$ , and  $\bar{y}_1, \dots, \bar{y}_n$  is a solution of  $\Sigma$ , analytic at  $x_0$ , then  $x_0$  is a point of contact between  $\bar{y}_1, \dots, \bar{y}_n$  and  $\mathfrak{N}$ .*

We prove Theorem 6 by using Theorem 5 in the same manner in which the lemma of §2 was used in proving Theorem 2. We simply replace the concept of a solution of  $\Sigma$  for which  $B$  is different from zero at  $x_0$ , by that of a solution in  $\mathcal{N}$ .

11. Extending this result in a special case, we assert that when  $\mathcal{F}$  consists purely of constants, then for every solution  $y_1(x), \dots, y_n(x)$  of  $\Sigma$ , analytic in an open region  $\mathfrak{B}$ , every point of  $\mathfrak{B}$  is a point of contact between  $y_1(x), \dots, y_n(x)$  and  $\mathcal{N}$ .

For every point  $x_1$  in  $\mathfrak{B}$ , every positive integer  $m$ , every  $\epsilon > 0$ , and every  $\delta > 0$ , there is a solution  $\bar{y}_1(x), \dots, \bar{y}_n(x)$  in  $\mathcal{N}$ , analytic at a point  $x_0$  in  $\mathfrak{B}$ , with  $|x_1 - x_0| < \delta$ , and with

$$|\bar{y}_{ij}(x_0) - y_{ij}(x_0)| < \epsilon, \quad i = 1, \dots, n; j = 0, 1, \dots, m.$$

We assume that  $\delta$  is sufficiently small so that

$$|y_{ij}(x_1) - y_{ij}(x_0)| < \epsilon, \quad i = 1, \dots, n; j = 0, 1, \dots, m.$$

Then  $|\bar{y}_{ij}(x_0) - y_{ij}(x_1)| < 2\epsilon$ , and this may be written  $|z_{ij}(x_1) - y_{ij}(x_1)| < 2\epsilon$ , where

$$z_i(x) \equiv \bar{y}_i(x + x_0 - x_1), \quad i = 1, \dots, n; j = 0, 1, \dots, m.$$

But  $z_1(x), \dots, z_n(x)$  is in  $\mathcal{N}$ , since  $\mathcal{F}$  consists purely of constants. Therefore  $x_1$  is a point of contact between  $y_1(x), \dots, y_n(x)$  and  $\mathcal{N}$ .

### PART III. EXAMPLES

12. We give an example of a sequence  $\Sigma_1, \Sigma_2, \dots$  and a solution  $y_1(x), \dots, y_n(x)$  of  $\Sigma$ , analytic in an open region  $\mathfrak{B}$ , such that the set of those points in  $\mathfrak{B}$  which are not points of contact between  $y_1(x), \dots, y_n(x)$  and  $\mathcal{N}$  is dense in  $\mathfrak{B}$ . From §11 we know that in such an example there must be functions in  $\mathcal{F}$  which are not constants.

Let  $\mathcal{F}$  consist of all rational functions of  $x$ . Let

$$(37) \quad a_1, a_2, \dots$$

be a sequence of points dense in the complex plane. Let  $\Sigma_1, \Sigma_2, \dots$  be the closed systems in one unknown  $y$  such that the manifold of  $\Sigma_n$ , ( $n = 1, 2, \dots$ ), is the family of functions

$$(38) \quad y = \frac{c_1}{x - a_1} + \frac{c_2}{(x - a_1)^2(x - a_2)} + \dots$$

$$+ \frac{c_n}{(x - a_1)^n(x - a_2)^{n-1} \dots (x - a_n)},$$

where the  $c_i$  are arbitrary constants.

Then it is easy to see that the manifold of  $\Sigma_n$  is identical with the manifold of a linear differential equation in  $y$ , with coefficients in  $\mathcal{F}$ . This equation affords a basic set for  $\Sigma_n$ ; therefore  $\Sigma_n$  is irreducible.\* Obviously the manifold of  $\Sigma_n$  is a proper part of the manifold of  $\Sigma_{n+1}$ .

Now  $\Sigma$  must have a non-empty set of arbitrary unknowns, as we have seen, so that  $y$  is a set of arbitrary unknowns for  $\Sigma$ . In other words, there is no nonzero form in  $\Sigma$ ; so every analytic function is a solution of  $\Sigma$ . Let  $\mathcal{N}$  be the union of the manifolds of the  $\Sigma_i$ ; that is, let  $\mathcal{N}$  be the union of the families of functions (38), ( $n=1, 2, \dots$ ). Let  $f(x)$  be a function which is analytic in an open region  $\mathfrak{B}$  and which is not in  $\mathcal{N}$ . Then the set of those points of the sequence (37) that lie in  $\mathfrak{B}$  is dense in  $\mathfrak{B}$ . No point in (37) is a point of contact between  $f(x)$  and  $\mathcal{N}$ , since for every positive integer  $l$  the only functions in  $\mathcal{N}$  which are analytic at the point  $a_l$  are the functions which are in the family (38) when  $n=l-1$ . Hence the complement in  $\mathfrak{B}$  of the set of points of contact between  $f(x)$  and  $\mathcal{N}$  is dense in  $\mathfrak{B}$ .

We note that there is no open subregion  $\mathfrak{B}_1$  of  $\mathfrak{B}$  in which  $f(x)$  may be approximated uniformly, with arbitrary closeness, by a solution in  $\mathcal{N}$ . For if such an open region  $\mathfrak{B}_1$  exists, then every point of  $\mathfrak{B}_1$  is a point of contact between  $f(x)$  and  $\mathcal{N}$ .

13. The phenomenon exemplified in the preceding section is in marked contrast with that appearing in the following example:

Let  $\Sigma_k$  be the closed irreducible† system in the unknown  $y$  with a basic set  $y_k, \ddagger$

$\Sigma$  is trivial as in the preceding example. Here, however, if  $f(x)$  is any function, analytic in an open region  $\mathfrak{B}$ , then every point of  $\mathfrak{B}$  has a neighborhood in which  $f(x)$  may be uniformly approximated by solutions of the  $\Sigma_i$ . In short, every polynomial is a solution of some system  $\Sigma_i$ .

14. In the example of §12, for certain solutions of  $\Sigma$  there existed no region in which uniform approximation by solutions of the  $\Sigma_i$  was possible. Each solution of  $\Sigma$  having this property had the additional property that every subregion of its domain of analyticity contained a point which was not a point of contact between the given solution and the union of the manifolds of the  $\Sigma_i$ . This second property of course implies the first. We shall prove that the converse is not true. That is, we shall exhibit a sequence  $\Sigma_1, \Sigma_2, \dots$  and a solution of the corresponding system  $\Sigma$ , such that every point in the domain of analyticity of the given solution is a point of contact between the solution and the union of the manifolds of the  $\Sigma_i$ , while on the other hand, there is

\* Cf. ADE, §§65, 45.

† The existence of such a  $\Sigma_k$  follows from two theorems of Ritt, ADE, §§65, 45.

‡ Subscripts indicate differentiation.

no open region in which the solution can be uniformly approximated by solutions of the  $\Sigma_i$ .

Let

$$(39) \quad \alpha_1, \alpha_2, \dots$$

be a sequence of complex constants. In terms of the sequence (39) we define a sequence of operators

$$(40) \quad \theta_1, \theta_2, \dots$$

as follows:

$$(41) \quad \theta_i g(x) = g(x)(g'(x) + \alpha_i g(x)), \theta_i A = A(A' + \alpha_i A), \quad i = 1, 2, \dots,$$

for every analytic function  $g(x)$  and every form  $A$ . Set  $\phi_k = \theta_k \theta_{k-1} \dots \theta_2 \theta_1$ , ( $k = 1, 2, \dots$ ). Let

$$(42) \quad \Sigma_1, \Sigma_2, \dots$$

be the sequence of closed systems such that the manifold of  $\Sigma_n$ , ( $n = 1, 2, \dots$ ), is the family of all functions  $y(x)$  which satisfy the equation

$$(43) \quad \frac{d}{dx} (\phi_n y(x)) = 0.$$

Set  $A_k \equiv \phi_k y$ , set  $B_k \equiv A'_k$ , and let  $S_k$  be the separant of  $B_k$ , ( $k = 1, 2, \dots$ ). Evidently the manifold of  $B_{i+1}$  includes the manifold of  $B_i$ , so that  $\Sigma_{i+1}$  holds  $\Sigma_i$ , ( $i = 1, 2, \dots$ ). Since  $S_i S_{i+1}$  does not hold  $B_i$ , the general solution of  $B_{i+1}$  includes the general solution of  $B_i$ .

We shall prove that  $\Sigma_n$  is irreducible. This is equivalent to proving that the manifold of  $B_n$  is identical with the general solution of  $B_n$ . It suffices to prove that the manifold of  $S_n$  is in the general solution of  $B_n$ . This last is easy to see when  $n = 1$ . We assume that it is true when  $n = r$ . Then the manifold of  $B_r$  is identical with the general solution of  $B_r$ . Now  $S_{r+1} = A_r S_r$ . Hence every solution of  $S_{r+1}$  is in the general solution of  $B_r$ , and consequently in the general solution of  $B_{r+1}$ .

Thus  $\Sigma_n$ , ( $n = 1, 2, \dots$ ), is irreducible. As in §12, the system  $\Sigma$  is the trivial system of which every analytic function is a solution.

Let  $\mathcal{M}_i$  be the manifold of  $\Sigma_i$ , ( $i = 1, 2, \dots$ ), and let  $\mathcal{N} = \mathcal{M}_1 + \mathcal{M}_2 + \dots$ .

Let  $y(x)$  be an analytic function such that for some  $k$  the function  $\phi_k y(x)$  is not identically zero, and has a zero at a point  $x_0$ . Then  $y(x)$  is not in  $\mathcal{N}$ . For suppose  $y(x)$  is in  $\mathcal{N}$ . There exists an  $n$  such that  $y(x)$  satisfies (43). We may and do assume that  $n > k$ . Let  $x_0$  be a zero of order  $\sigma$  for  $\phi_k y(x)$ . Then it is easy to see that  $x_0$  is a zero of order  $2^{n-k}(\sigma-1)+1$  for  $\phi_n y(x)$ . But this is

impossible, since  $\phi_n y(x)$  is a constant, by equation (43). Hence  $y(x)$  is not in  $\mathcal{N}$ .

Let  $f(x)$  be a polynomial of positive degree. Then for every choice of a sequence (39) the function  $\phi_k f(x)$  is a polynomial of positive degree ( $k=1, 2, \dots$ ). Let  $\mathfrak{R}_1, \mathfrak{R}_2, \dots$  be a sequence of open sets such that for every open region  $\mathfrak{B}$  there is a  $k$  such that  $\mathfrak{R}_k$  is included in  $\mathfrak{B}$ . We shall choose a sequence (39) in such a way that for every  $k$  the function  $\phi_k f(x)$  has a zero in  $\mathfrak{R}_k$ .

We take a point  $b_1$  in  $\mathfrak{R}_1$  such that  $f(b_1) \neq 0$ , and define  $\alpha_1 = -f'(b_1)/f(b_1)$ . Then  $\phi_1 f(x)$  has the zero  $b_1$  in  $\mathfrak{R}_1$ . When  $\alpha_1, \alpha_2, \dots, \alpha_{n-1}$  have been chosen so that  $\phi_k f(x)$  has a zero  $b_k$  in  $\mathfrak{R}_k$ , ( $k=1, 2, \dots, n-1$ ), we take a point  $b_n$  in  $\mathfrak{R}_n$  at which  $\phi_{n-1} f(x)$  is not zero, and, letting  $g(x) \equiv \phi_{n-1} f(x)$ , we define  $\alpha_n = -g'(b_n)/g(b_n)$ . Then  $\phi_n f(x)$  has the zero  $b_n$  in  $\mathfrak{R}_n$ . Proceeding in this manner we determine a sequence of points

$$(44) \quad b_1, b_2, \dots,$$

dense in the complex plane, and a choice of the sequence (39), such that the point  $b_k$  is a zero of  $\phi_k f(x)$ , ( $k=1, 2, \dots$ ). We now retain this choice of (39).

Now suppose that there exists an open region  $\mathfrak{B}_1$  in which  $f(x)$  can be approximated uniformly by solutions of the  $\Sigma_i$ . Let  $y_1(x), y_2(x), \dots$  be a sequence of functions in  $\mathcal{N}$ , converging to  $f(x)$  uniformly in  $\mathfrak{B}_1$ . Then for each  $k$  the sequence  $\phi_k y_1(x), \phi_k y_2(x), \dots$  converges to  $\phi_k f(x)$  uniformly in  $\mathfrak{B}_1$ . Let  $k$  be such that  $\phi_k f(x)$  has a zero in  $\mathfrak{B}_1$ .<sup>\*</sup> Then there is an  $m$  such that  $\phi_k y_m(x)$  is not identically zero and has a zero in  $\mathfrak{B}_1$ . We have seen that this implies that  $y_m(x)$  is not in  $\mathcal{N}$ . This contradiction implies that there is no open region in which  $f(x)$  can be approximated uniformly by solutions of the  $\Sigma_i$ . On the other hand, every point of the complex plane is a point of contact between  $f(x)$  and  $\mathcal{N}$ , since  $\mathcal{F}$  is a field of constants.

Another such example can be constructed as follows: Let  $\theta$  be the operator such that

$$\theta g(x) = g(x)[2(g'(x))^3 - 3g(x)g'(x)g''(x) + (g(x))^2g'''(x)],$$

for every analytic function  $g(x)$ . Let  $\Sigma_1, \Sigma_2, \dots$  be the sequence of closed systems such that the manifold of  $\Sigma_n$ , ( $n=1, 2, \dots$ ), is the family of all functions  $y(x)$  which satisfy the equation

$$(45) \quad \frac{d}{dx}(\theta^n y(x)) = 0.$$

Arguments similar to those used in the preceding example show that  $\Sigma_i$  is

<sup>\*</sup> We are using here the fact that the sequence (44) is dense in the complex plane.

irreducible and is held by  $\Sigma_{i+1}$ , ( $i=1, 2, \dots$ ), and also show that if  $y(x)$  is a function such that  $\theta^k y(x)$  has a simple zero for some  $k$ , then  $y(x)$  is not in  $\mathcal{N}$ .

Let  $\omega_1, \omega_2$  be any two complex numbers whose ratio is not real, and let  $\sigma(x)$  be the corresponding  $\sigma$ -function of Weierstrass.\* Using a functional equation of  $\sigma(x)$ ,† we find that

$$\theta^k \sigma(x) = 2^{3k(k-1)/2} \sigma(2^k x), \quad k = 1, 2, \dots$$

Every circle which is of diameter greater than  $|\omega_1| + |\omega_2|$  contains a simple zero of  $\sigma(x)$ , so that every circle which is of diameter greater than  $(|\omega_1| + |\omega_2|)/2^k$  contains a simple zero of  $\theta^k \sigma(x)$ .

Thus if  $\sigma(x)$  can be approximated uniformly in an open region by solutions of the  $\Sigma_i$ , there is a  $k$  such that for some function  $y(x)$  in  $\mathcal{N}$  the function  $\theta^k y(x)$  has a simple zero at some point in that open region. This contradiction proves that there is no open region in which  $\sigma(x)$  can be approximated uniformly by solutions of the  $\Sigma_i$ .

15. Let  $\Sigma_1, \Sigma_2, \dots$  be a sequence of closed irreducible systems in  $y_1, \dots, y_n$  such that the manifold of  $\Sigma_i$  is a proper part of the manifold of  $\Sigma_{i+1}$ , ( $i=1, 2, \dots$ ). Let  $\Sigma$  be the set of all forms  $F$  such that  $F$  holds every  $\Sigma_i$ . Discarding a finite set of the  $\Sigma_i$ , if necessary, we assume that there is a fixed set of unknowns  $y_1, \dots, y_q$  which is a set of arbitrary unknowns for every  $\Sigma_i$ .‡ Of course  $q < n$ . Let  $A_{i,q+1}, A_{i,q+2}, \dots, A_{i,n}$  be a corresponding basic set for  $\Sigma_i$ , with  $A_{ij}$  introducing  $y_j$ , and let  $\gamma_{ij}$  be the order of  $A_{ij}$  in  $y_j$ , ( $j=q+1, \dots, n; i=1, 2, \dots$ ). Let  $r$  be the number of arbitrary unknowns in  $\Sigma$ . Then  $q < r \leq m$ . If  $\Sigma$  is not trivial, that is, if  $r < n$ , let  $A_{r+1}, A_{r+2}, \dots, A_n$  be a basic set for  $\Sigma$ , with  $A_k$  introducing  $y_k$ ,§ and let  $\gamma_k$  be the order of  $A_k$  in  $y_k$ , ( $k=r+1, \dots, n$ ). We consider the question of whether the values of  $r$ , and (when  $\gamma_k$  exist) of the  $\gamma_k$ , are determined uniquely by  $n, q$ , and the  $\gamma_{ij}$ . We know that the answer is affirmative when  $n-q=1$ , since in this case  $n-r=0$  and  $\Sigma$  is trivial. We shall indicate by examples in §§15, 16 that the answer is negative when  $n-q \geq 2$ .

Let  $\sigma$  be any nonnegative integer, and let  $\Sigma_n$ , ( $n=1, 2, \dots$ ), be the closed irreducible system in  $u, y$  with a basic set  $A_n, B_n$  where

$$A_n \equiv u_n, \quad B_n \equiv y - \sum_{i=1}^{n-1} i^\sigma u_i. ||$$

Since the order of  $A_n$  in  $u$  becomes infinite with  $n$ ,  $\Sigma$  cannot have a non-

\* Hurwitz and Courant, *Funktionentheorie*, p. 179.

† Ibid., p. 184.

‡ We renumber the unknowns, if necessary.

§ We renumber the unknowns again, if necessary.

|| For the existence of such a  $\Sigma_n$  cf. ADE, §§65, 45.

zero form in  $u$  alone. We shall prove that  $\Sigma$  has a nonzero form in  $u$  and  $y$ , so that  $u$  is a set of arbitrary unknowns for  $\Sigma$ . Moreover, we shall prove that the nonzero forms in  $\Sigma$  of lowest order in  $y$  are of order  $\sigma+1$  in  $y$ . Set

$$C_{nj} \equiv y_j - \sum_{i=1}^{n-j-1} i^{\sigma} u_{i+j}, \quad j = 0, 1, \dots, \sigma+1; n = 1, 2, \dots$$

Evidently  $C_{nj}$  is a form in  $\Sigma_n$  since it is a linear combination of derivatives of  $A_n$  and  $B_n$ . Set

$$D_n \equiv \sum_{j=0}^{\sigma+1} (-1)^j {}_{\sigma+1}C_j C_{nj}.*$$

It is easy to see that

$$D_n \equiv \sum_{j=0}^{\sigma+1} (-1)^j {}_{\sigma+1}C_j \left[ y_j - \sum_{k=j+1}^{\sigma+2} (k-j)^{\sigma} u_k \right]. \dagger$$

Since  $D_n$  is independent of  $n$ ,  $D_n$  is in  $\Sigma_j$  for every  $j$ . Therefore  $\Sigma$  contains the nonzero form  $D_n$  which is of order  $\sigma+1$  in  $y$ . Let  $m$  be any positive integer greater than  $\sigma-1$ , and let  $a_0, a_1, \dots, a_m; b_0, b_1, \dots, b_{\sigma}$  be any set of complex numbers. Let  $n=m+\sigma+2$ . We shall prove that  $\Sigma_n$  has a solution  $u(x), y(x)$  such that  $u_i(0) = a_i, y_k(0) = b_k, (i=0, 1, \dots, m; k=0, 1, \dots, \sigma)$ .

It is easily seen that such a solution exists if the system of linear equations

$$(46) \quad b_k = \sum_{i=1}^{m-k} i^{\sigma} a_{i+k} + \sum_{\alpha=m+1}^{n-1} (\alpha-k)^{\sigma} u_{\alpha}, \quad k = 0, 1, \dots, \sigma,$$

has a solution in  $u_{\alpha}, (\alpha=m+1, \dots, n-1)$ .

The system (46) will have a solution if the determinant  $d(m, \sigma)$  is not zero, where the element of  $d(m, \sigma)$  in the  $i$ th row and  $j$ th column is  $(m+1+j-i)^{\sigma}, (i, j=1, \dots, \sigma+1)$ . By repeated subtraction of adjacent columns, and then of adjacent rows, in  $d(m, \sigma)$ , it can be shown that  $d(m, \sigma) = (\sigma!)^{\sigma+1} \neq 0$ .

Hence  $\Sigma$  has a solution  $u(x), y(x)$  such that

$$u_i(0) = a_i, y_k(0) = b_k, \quad i = 0, 1, \dots, m; k = 0, 1, \dots, \sigma,$$

where  $m$  is any positive integer and the  $a_i$  and  $b_k$  are arbitrary. This obviously implies that  $\Sigma$  has no nonzero form whose order in  $y$  is less than  $\sigma+1$ . Evidently if  $u$  is taken as a set of arbitrary unknowns for  $\Sigma$ , then  $\sigma+1$  is the order in  $y$  of a basic set for  $\Sigma$ .

\* Note that  ${}_{\sigma+1}C_j$  is a binomial coefficient, not a form.

† This reduces quickly to proving the identity  $\sum_{j=0}^{\sigma+1} (-1)^j {}_{\sigma+1}C_j (k-j)^{\sigma} = 0$ , which holds for all complex  $k$ , since the left-hand member is the  $\sigma$ th derivative at the origin of  $e^{kx}(1-e^{-x})^{\sigma+1}$ .

16. Let  $\Sigma_n$ , ( $n=1, 2, \dots$ ), be the closed irreducible system in  $u, y$  with a basic set  $A_n, B_n$  where  $A_n \equiv u_n, B_n \equiv y - \sum_{i=1}^{n-1} i! u_i$ . We shall prove that for every positive integer  $m$  and every set of complex numbers  $a_0, a_1, \dots, a_m; b_0, b_1, \dots, b_m$  the system  $\Sigma_{2m+2}$  has a solution  $u, y$  in which  $u_i(0) = a_i, y_i(0) = b_i, (i, j=0, 1, \dots, m)$ . It suffices to prove that the system of linear equations

$$(47) \quad b_k = \sum_{i=1}^{m-k} i! a_{i+k} + \sum_{\alpha=m+1}^{2m+1} (\alpha - k)! u_\alpha, \quad k = 0, 1, \dots, m,$$

has a solution in  $u_\alpha, (\alpha=m+1, \dots, 2m+1)$ . Such a solution exists because the determinant  $d_m$  of the coefficients is not zero, where the element of  $d_m$  in the  $i$ th row and  $j$ th column is  $(m+1+j-i)!, (i, j=1, \dots, m+1)$ . The value of  $d_m$  is easily seen to be  $(-1)^{m(m+1)/2} (1!2! \dots m!)^2 (m+1)!$ .

Consequently, for every set of complex numbers  $a_0, a_1, \dots, a_m; b_0, b_1, \dots, b_m$ ,  $\Sigma$  has a solution  $u(x), y(x)$  with  $u_i(0) = a_i, y_i(0) = b_i, (i, j=0, 1, \dots, m)$ . This means that  $\Sigma$  has no nonzero form. That is,  $\Sigma$  is the trivial system having  $u, y$  as a set of arbitrary unknowns.

17. Let  $\Sigma$  be any closed irreducible system in  $u_1, \dots, u_q; y_1, \dots, y_p$ , where  $u_1, \dots, u_q, (q \geq 1)$ , is a set of arbitrary unknowns. Let  $s$  be any integer with  $0 \leq s < q$ . We assert that there is a sequence

$$(48) \quad \Sigma_1, \Sigma_2, \dots$$

of closed irreducible systems such that  $\Sigma_i$  has  $s$  arbitrary unknowns, such that the manifold of  $\Sigma_i$  is a proper part of the manifold of  $\Sigma_{i+1}, (i=1, 2, \dots)$ , and such that  $\Sigma$  is the totality of forms common to the  $\Sigma_i, (i=1, 2, \dots)$ .

For example, we may construct a sequence (48) as follows:

Let  $A_1, A_2, \dots, A_p$  be a basic set for  $\Sigma$ , with  $A_i$  introducing  $y_i$ . Let  $m$  be a positive integer greater than the order of  $A_i$  in  $u_i, (i=1, \dots, p; j=1, \dots, q-s)$ . Then for every positive integer  $n$  the ascending set

$$(49) \quad u_{1,m+n}, u_{2,m+n}, \dots, u_{q-s,m+n}, A_1, \dots, A_p$$

is the basic set of a closed irreducible system. This is a simple consequence of Ritt's theorems characterizing the basic set of a closed irreducible system.\* Let  $\Sigma_n$  be the closed irreducible system having (49) for a basic set ( $n=1, 2, \dots$ ). It is easy to see that with this definition of  $\Sigma_n$  the sequence (48) has the desired properties.

\* ADE, §§65, 45.

# ON THE COEFFICIENTS OF CERTAIN MODULAR FORMS BELONGING TO SUBGROUPS OF THE MODULAR GROUP\*

BY

HERBERT S. ZUCKERMAN†

1. **Introduction.** The Fourier coefficients of modular forms  $F(\tau)$  of positive dimension have been determined‡ for the case in which  $F(\tau)$  belongs to the full modular group. It is the purpose of this paper to generalize some of those results to the case where  $F(\tau)$  belongs to an arbitrary subgroup of the modular group subject to certain restrictions.

By  $F(\tau)$  belonging to a subgroup  $\gamma$  we mean that  $F(\tau)$  is analytic in the upper half-plane  $I(\tau) > 0$  and that  $F(\tau)$  satisfies a transformation equation

$$(1.11) \quad F(\tau') = \epsilon(-i(c\tau + d))^{-r} F(\tau)$$

for every transformation

$$(1.12) \quad \tau' = \frac{a\tau + b}{c\tau + d}$$

of  $\gamma$ . In (1.11)  $r$ , which we shall assume throughout to be positive, is the dimension of  $F(\tau)$  and  $\epsilon = \epsilon(a, b, c, d)$  is of absolute value one and depends only on the transformation (1.12). If  $c \neq 0$  we take  $c > 0$ , assign

$$(1.13) \quad -\frac{\pi}{2} < \arg(-i(c\tau + d)) < \frac{\pi}{2},$$

and define  $(-i(c\tau + d))^{-r}$  as  $|c\tau + d|^{-r} \exp\{-ir \arg(-i(c\tau + d))\}$ .

In the case of the full modular group, a Fourier expansion for  $F(\tau)$  was found by considering a parabolic transformation which had infinity as fixed point. In the case of a subgroup  $\gamma$  we must consider a set of parabolic transformations such that no two of the fixed points of the transformations are equivalent under  $\gamma$ . For these fixed points we take parabolic vertices of a fundamental region of  $\gamma$ . The expansions corresponding to the point at infinity are simpler than those corresponding to finite points. However, in general, the fundamental region will have more than one parabolic point; so

\* Presented to the Society, February 26, 1938; received by the editors February 17, 1938.

† National Research Fellow.

‡ H. Rademacher and H. S. Zuckerman, *On the Fourier coefficients of certain modular forms of positive dimension*, *Annals of Mathematics*, (2), vol. 39 (1938), pp. 433-462.

some finite points will have to be considered. In order to be able to treat all the expansions symmetrically, we choose a fundamental region which does not have the point at infinity as a vertex. We then show, in §3, that  $F(\tau)$  has a set of expansions of the form

$$(1.2) \quad F(\tau) = (-i(\tau - P_\theta))^r x^{\alpha_\theta} \sum_m a_m^{(\theta)} x^m = (-i(\tau - P_\theta))^r x^{\alpha_\theta} f_\theta(x),$$

$$x = \exp \{ -2\pi i/c_\theta(\tau - P_\theta) \}, \quad g = 1, 2, \dots, s,$$

corresponding to the parabolic points  $P_\theta$ .

In order that  $s$  in (1.2) be finite we assume that  $\gamma$  is of finite index in the full modular group. This is the only restriction that we place on  $\gamma$ . We make use of the fact that the subgroup is defined by its fundamental region and hence do not need any arithmetical characterization of it. That is, the existence in  $\gamma$  of all the transformations which we use is an immediate consequence of the form of the fundamental region.

To the conditions that  $r$  is positive and that  $F(\tau)$  is analytic for  $I(\tau) > 0$  we add the restriction that  $F(\tau)$  shall have only polar singularities, measured in the uniformizing variable  $x$  of (1.2), at the points  $P_\theta$ . That is, we assume that only a finite number of  $a_m^{(\theta)}$  with negative  $m$  are different from zero. In §§5 and 6 we find expressions for the  $a_m^{(\theta)}$  for  $m > 0$  in terms of the  $a_m^{(\theta)}$  for  $m < 0$ .

The expansions (1.2) which we obtain are Fourier expansions in the variable  $(\tau - P_\theta)^{-1}$ . These expansions can be transformed into usual Fourier expansions in the variable  $\tau$ . We consider any transformation of the full modular group

$$(1.31) \quad \tau' = \frac{A\tau + B}{C\tau + D}$$

and write

$$(1.32) \quad F(\tau') = \epsilon^* (-i(C\tau + D))^{-r} F^*(\tau)$$

where  $F^*(\tau)$  depends on the choice of the transformation (1.31) as well as on  $\tau$ . In §3 we find an expression, (3.23), for  $F^*(\tau)$  in terms of the  $f_\theta(x)$  which is, in fact, a usual Fourier expansion in  $\tau$  since, by definition, we have

$$f_\theta(x) = \sum_m a_m^{(\theta)} x^m.$$

In particular, we can obtain a usual Fourier expansion for  $F(\tau)$  itself by choosing the identity transformation for (1.31).

In §7 we specialize the results of §§5 and 6 to the case of a particular subgroup, evaluating all the constants which depend only on the choice of  $\gamma$ . In §8 we consider the function  $\vartheta_2(0|\tau)^{-1}$  which belongs to the subgroup of §7. From the  $F^*(\tau)$  of (1.32) we then get expansions for  $\vartheta_2(0|\tau)^{-1}$ ,  $\vartheta_3(0|\tau)^{-1}$ , and  $\vartheta_4(0|\tau)^{-1}$ . This particular function is considered because it was partially treated by Hardy and Ramanujan.<sup>†</sup> They considered only the expansion of  $\vartheta_4(0|\tau)^{-1}$  and obtained results which compare with ours in the same way as their results for the partition function compare with those obtained by Rademacher.<sup>‡</sup>

**2. The fundamental region.** We choose a fundamental region  $R$  of  $\gamma$  which we shall keep fixed throughout the discussion. Although we could use any fundamental region, we find it convenient, for symmetry, to choose one that does not have the point at infinity as a vertex. This choice is clearly possible since the subgroup  $\gamma$  is of finite index. Such a region  $R$  can be obtained from a fundamental region  $R'$  of  $\gamma$ , which has the point at infinity as a vertex, by subjecting  $R'$  to a transformation of  $\gamma$  which takes the point at infinity into a finite point and which does not take any of the finite vertices of  $R'$  into infinity.

Under  $\gamma$  every real rational point transforms into a real rational point or infinity. Since every real rational point is equivalent to a point of  $R$ , we see that each is equivalent to a real rational point of  $R$ . Also every real point of  $R$  is a parabolic vertex; so we see that every real rational point is equivalent under  $\gamma$  to a parabolic vertex of  $R$ .

Some of the parabolic vertices of  $R$  may be equivalent under  $\gamma$  thus forming cycles of more than one vertex. We pick out a representative vertex from each cycle and name the vertices  $P_1, P_2, \dots, P_s$ . Since the  $P_g$  are real rational points, we can write

$$(2.1) \quad P_g = \frac{p_g}{q_g}, \quad (p_g, q_g) = 1, \quad q_g > 0, \quad g = 1, 2, \dots, s.$$

**3. The functions  $f_g(x)$ .** Corresponding to each parabolic vertex  $P_g$  there is a transformation of  $\gamma$  which may be written

$$(3.11) \quad \frac{1}{\tau' - P_g} = \frac{1}{\tau - P_g} + c_g, \quad c_g > 0.$$

Putting this transformation in our standard form, we have

<sup>†</sup> *Asymptotic formulae in combinatorial analysis*, Proceedings of the London Mathematical Society, (2), vol. 17 (1918), pp. 75-115; also S. Ramanujan, *Collected Papers*, pp. 276-309.

<sup>‡</sup> *On the partition function  $p(n)$* , Proceedings of the London Mathematical Society, (2), vol. 43 (1937), pp. 241-254.

$$\tau' = \frac{\delta^{-1}(q_\theta c_\theta p_\theta + q_\theta^2)\tau - \delta^{-1}c_\theta p_\theta^2}{\delta^{-1}q_\theta^2 c_\theta \tau + \delta^{-1}(q_\theta^2 - c_\theta p_\theta q_\theta)},$$

where  $\delta = (q_\theta c_\theta p_\theta + q_\theta^2, c_\theta p_\theta^2, q_\theta^2 c_\theta, q_\theta^2 - c_\theta p_\theta q_\theta) = (q_\theta^2, c_\theta)$ . However this is a modular transformation; so its determinant is unity. Hence we have

$$q_\theta^4 - c_\theta^2 p_\theta^2 q_\theta^2 + c_\theta^2 p_\theta^2 q_\theta^2 = \delta^2, \quad \delta = \pm q_\theta^2;$$

so  $q_\theta^2$  divides  $c_\theta$  and we may write the transformation in our standard form as

$$\tau' = \frac{(c_\theta P_\theta + 1)\tau - c_\theta P_\theta^2}{c_\theta \tau + 1 - c_\theta P_\theta}.$$

Then, by (1.11), we have

$$F(\tau') = \epsilon(-i(c_\theta \tau + 1 - c_\theta P_\theta))^{-r} F(\tau),$$

or, writing  $T = (\tau - P_\theta)^{-1}$ ,  $T' = (\tau' - P_\theta)^{-1} = T + c_\theta$ ,

$$\begin{aligned} F\left(\frac{1}{T'} + P_\theta\right) &= \epsilon\left(-i\left(\frac{c_\theta}{T} + 1\right)\right)^{-r} F\left(\frac{1}{T} + P_\theta\right), \\ F\left(\frac{1}{T + c_\theta} + P_\theta\right) &= \epsilon e^{\pi i r/2} \frac{(i(T + c_\theta))^{-r}}{(iT)^{-r}} F\left(\frac{1}{T} + P_\theta\right). \end{aligned}$$

Defining  $\alpha_\theta$  by  $\epsilon e^{\pi i r/2} = e^{-2\pi i \alpha_\theta}$ , ( $0 \leq \alpha_\theta < 1$ ), we have

$$\begin{aligned} \exp\left\{\frac{2\pi i \alpha_\theta}{c_\theta}(T + c_\theta)\right\} (iT + c_\theta)^r F\left(\frac{1}{T + c_\theta} + P_\theta\right) \\ = \exp\left\{\frac{2\pi i \alpha_\theta}{c_\theta}T\right\} (iT)^r F\left(\frac{1}{T} + P_\theta\right) \end{aligned}$$

and therefore obtain a Fourier expansion

$$\exp\left\{\frac{2\pi i \alpha_\theta}{c_\theta}T\right\} (iT)^r F\left(\frac{1}{T} + P_\theta\right) = \sum_m a_m^{(\theta)} \exp\left\{-\frac{2\pi i m T}{c_\theta}\right\}$$

or, going back to the variable  $\tau$ ,

$$\begin{aligned} F(\tau) &= \exp\left\{-\frac{2\pi i \alpha_\theta}{c_\theta} \frac{1}{\tau - P_\theta}\right\} \left(\frac{i}{\tau - P_\theta}\right)^{-r} \sum_m a_m^{(\theta)} \exp\left\{-\frac{2\pi i m}{c_\theta} \frac{1}{\tau - P_\theta}\right\} \\ &= x^{\alpha_\theta} (-i(\tau - P_\theta))^r \sum_m a_m^{(\theta)} x^m \end{aligned}$$

with  $x = \exp\left\{-2\pi i/(c_\theta(\tau - P_\theta))\right\}$ . We place the restriction on  $F(\tau)$  that only a finite number  $\mu_\theta$  of terms have negative exponents and write

$$\begin{aligned}
 (3.12) \quad F(\tau) &= (-i(\tau - P_g))^{r_g} \sum_{m=\mu_g}^{\infty} a_m^{(g)} x^m = (-i(\tau - P_g))^{r_g} f_g(x), \\
 x &= \exp \left\{ -\frac{2\pi i}{c_g(\tau - P_g)} \right\}, \quad g = 1, 2, \dots, s.
 \end{aligned}$$

The functions  $f_g(x)$  are regular inside the unit circle except, when  $\mu_g > 0$ , for poles of order  $\mu_g$  at  $x=0$ . The functions

$$(3.13) \quad P_g(x) = a_{-\mu_g}^{(g)} x^{-\mu_g} + \dots + a_{-1}^{(g)} x^{-1}, \quad g = 1, 2, \dots, s,$$

are the principal parts of the  $f_g(x)$  at  $x=0$ . We understand  $P_g(x)$  to be zero if  $\mu_g=0$ .

In the following sections we shall determine the  $a_m^{(g)}$  in terms of the constants of (1.11) and (3.13). The functions  $F^*(\tau)$  of (1.32) are then found as follows. We consider the point  $A/C$  where  $A$  and  $C$  are the coefficients of  $\tau$  in (1.31). If  $C \neq 0$  we take  $C > 0$  and the point is a real rational point. If  $C=0$  the point is the point at infinity. In either case  $A/C$  is congruent to some parabolic vertex  $P_l$  of  $R$ , where  $l$  is determined by  $A/C$  and hence by the transformation. Therefore we can find a transformation

$$(3.21) \quad \tau' = \frac{a_1\tau + b_1}{c_1\tau + d_1}, \quad c_1 \geq 0,$$

of  $\gamma$  which takes  $P_g$  into the point  $A/C$ . That is

$$\frac{a_1 p_g + b_1 q_g}{c_1 p_g + d_1 q_g} = \frac{A}{C};$$

hence

$$(3.22) \quad a_1 p_g + b_1 q_g = \kappa A, \quad c_1 p_g + d_1 q_g = \kappa C.$$

Solving these equations for  $p_g$  and  $q_g$ , we have

$$p_g = \kappa(A d_1 - C b_1), \quad q_g = \kappa(C a_1 - A c_1);$$

then, since  $(p_g, q_g) = 1$ , we have  $\kappa = \pm 1$ . Using these equations and the fact that the determinant of (3.21) is unity, we find

$$\frac{A\tau + B}{C\tau + D} = \frac{a_1 \left( \frac{\kappa p_g \tau + d_1 B - b_1 D}{\kappa q_g \tau + a_1 D - c_1 B} \right) + b_1}{c_1 \left( \frac{\kappa p_g \tau + d_1 B - b_1 D}{\kappa q_g \tau + a_1 D - c_1 B} \right) + d_1};$$

then, by (1.11), we obtain the relation

$$F\left(\frac{A\tau + B}{C\tau + D}\right) = \epsilon(a_1, b_1, c_1, d_1) \left( -i \left( c_1 \left( \frac{\kappa p_\sigma \tau + d_1 B - b_1 D}{\kappa q_\sigma \tau + a_1 D - c_1 B} \right) + d_1 \right) \right)^{-r} \\ \cdot F\left(\frac{\kappa p_\sigma \tau + d_1 B - b_1 D}{\kappa q_\sigma \tau + a_1 D - c_1 B}\right).$$

Now applying (3.12) to the function on the right and simplifying, we obtain

$$F\left(\frac{A\tau + B}{C\tau + D}\right) = \epsilon(a_1, b_1, c_1, d_1) \exp \left\{ \kappa \frac{\pi i r}{2} \right\} \exp \left\{ \kappa \frac{2\pi i q_\sigma}{c_\sigma} \alpha_\sigma (a_1 D - c_1 B) \right\} \\ \cdot (-i(C\tau + D))^{-r} q_\sigma^{-r} \exp \left\{ \frac{2\pi i q_\sigma^2}{c_\sigma} \alpha_\sigma \tau \right\} \\ \cdot \sum_{m=-\mu_\sigma}^{\infty} a_m^{(\sigma)} \exp \left\{ \kappa \frac{2\pi i q_\sigma}{c_\sigma} m (a_1 D - c_1 B) \right\} \exp \left\{ \frac{2\pi i q_\sigma^2}{c_\sigma} m \tau \right\} \\ = \epsilon(a_1, b_1, c_1, d_1) \exp \left\{ \kappa \frac{\pi i r}{2} \right\} \exp \left\{ \kappa \frac{2\pi i q_\sigma}{c_\sigma} \alpha_\sigma (a_1 D - c_1 B) \right\} \\ \cdot (-i(C\tau + D))^{-r} q_\sigma^{-r} \exp \left\{ \frac{2\pi i q_\sigma^2}{c_\sigma} \alpha_\sigma \tau \right\} \\ \cdot f_\sigma \left( \exp \left\{ \kappa \frac{2\pi i q_\sigma}{c_\sigma} (a_1 D - c_1 B) \right\} \exp \left\{ \frac{2\pi i q_\sigma^2}{c_\sigma} \tau \right\} \right);$$

therefore we have

$$(3.23) \quad F^*(\tau) = q_\sigma^{-r} \exp \left\{ \frac{2\pi i q_\sigma^2}{c_\sigma} \alpha_\sigma \tau \right\} \\ \cdot f_\sigma \left( \exp \left\{ \kappa \frac{2\pi i q_\sigma}{c_\sigma} (a_1 D - c_1 B) \right\} \exp \left\{ \frac{2\pi i q_\sigma^2}{c_\sigma} \tau \right\} \right),$$

$$(3.24) \quad \epsilon^* = \epsilon(a_1, b_1, c_1, d_1) \exp \left\{ \kappa \frac{\pi i r}{2} \right\} \exp \left\{ \kappa \frac{2\pi i q_\sigma}{c_\sigma} \alpha_\sigma (a_1 D - c_1 B) \right\}.$$

4. **The transformation equation.** In order to make use of the transformation equation (1.11) in the next section we shall need it in a special form. We consider all rational real nonnegative numbers  $h/k$ ,  $(h, k)=1$ ,  $k>0$ ,  $h \geq 0$ . Then by (3.12) we have, with  $\tau = P_\sigma - k/(c_\sigma(iz+h))$ ,

$$(4.11) \quad f_\sigma \left( \exp \left\{ 2\pi i \left( \frac{h}{k} + \frac{iz}{k} \right) \right\} \right) = \exp \left\{ -2\pi i \alpha_\sigma \left( \frac{h}{k} + \frac{iz}{k} \right) \right\} \\ \cdot \left( \frac{k}{c_\sigma(z-ih)} \right)^{-r} F \left( P_\sigma - \frac{k}{c_\sigma(iz+h)} \right).$$

Now  $P_\theta - k/c_\theta h$  is a real rational point and hence is congruent under  $\gamma$  to some parabolic vertex  $P_\beta$  of  $R$  where  $\beta = \beta(h, k, g)$ . Then there is a transformation of  $\gamma$ ,

$$\tau' = \frac{a\tau + b}{c\tau + d},$$

where  $a, b, c$ , and  $d$  are determined, although not uniquely, by  $h, k$ , and  $g$ , which takes  $P_\theta - k/c_\theta h$  into  $P_\beta$ , that is,

$$(4.12) \quad \frac{p_\beta}{q_\beta} = \left\{ a \left( \frac{p_\theta}{q_\theta} - \frac{k}{c_\theta h} \right) + b \right\} \left\{ c \left( \frac{p_\theta}{q_\theta} - \frac{k}{c_\theta h} \right) + d \right\}^{-1}.$$

We have seen that  $q_\theta^2$  divides  $c_\theta$ , so we have

$$(4.13) \quad \begin{aligned} a \left( \frac{c_\theta}{q_\theta} p_\theta h - k \right) + bc_\theta h &= \sigma_{h,k}^{(\theta)} p_\beta, \\ c \left( \frac{c_\theta}{q_\theta} p_\theta h - k \right) + dc_\theta h &= \sigma_{h,k}^{(\theta)} q_\beta. \end{aligned}$$

From these equations and the fact that  $h$  and  $k$  are relatively prime, it follows that  $\sigma_{h,k}^{(\theta)}$  divides  $c_\theta$ , which implies

$$(4.14) \quad -c_\theta \leq \sigma_{h,k}^{(\theta)} \leq c_\theta, \quad \sigma_{h,k}^{(\theta)} \neq 0.$$

Using this transformation we have by (1.11)

$$(4.15) \quad \begin{aligned} F \left( P_\theta - \frac{k}{c_\theta(iz + h)} \right) &= (\epsilon_{h,k}^{(\theta)})^{-1} \left( -i \left( cP_\theta - \frac{ck}{c_\theta(iz + h)} + d \right) \right)^r \\ &\quad \cdot F \left[ \frac{a \left( P_\theta - \frac{k}{c_\theta(iz + h)} \right) + b}{c \left( P_\theta - \frac{k}{c_\theta(iz + h)} \right) + d} \right] \\ &= (\epsilon_{h,k}^{(\theta)})^{-1} \left( -i \frac{c_\theta(cP_\theta + d)iz + \sigma_{h,k}^{(\theta)} q_\beta}{c_\theta(iz + h)} \right)^r \\ &\quad \cdot F \left( P_\beta + \frac{kc_\theta iz}{\sigma_{h,k}^{(\theta)} q_\beta (c_\theta(cP_\theta + d)iz + \sigma_{h,k}^{(\theta)} q_\beta)} \right), \end{aligned}$$

where  $\epsilon_{h,k}^{(\theta)} = \epsilon(a, b, c, d)$ , the  $\epsilon$  of (1.11) associated with the transformation determined by (4.12). We now apply the Fourier expansion (3.12) to the function on the right in (4.15), combine it with (4.11), and obtain after some simplification

$$\begin{aligned}
 f_{\theta} \left( \exp \left\{ 2\pi i \left( \frac{h}{k} + iz \right) \right\} \right) \\
 = (\epsilon_{h,k}^{(\theta)})^{-1} \exp \left\{ \left( 1 - \delta_{h,k}^{(\theta)} \right) \frac{\pi i r}{2} \right\} \\
 \cdot \exp \left\{ - \frac{2\pi i}{k} \left( \alpha_{\theta} h + \frac{\sigma_{h,k}^{(\theta)} q_{\beta} (cP_{\theta} + d) \alpha_{\beta}}{c_{\beta}} \right) \right\} \\
 \cdot \left( \frac{c_{\theta} z}{|\sigma_{h,k}^{(\theta)}| q_{\beta}} \right)^r \exp \left\{ \frac{2\pi}{k} \left( \alpha_{\theta} z - \frac{\alpha_{\beta} (\sigma_{h,k}^{(\theta)})^2 q_{\beta}^2}{c_{\beta} c_{\theta}} \cdot \frac{1}{z} \right) \right\} \\
 \cdot f_{\beta} \left( \exp \left\{ - \frac{2\pi}{k c_{\beta}} \left( \sigma_{h,k}^{(\theta)} q_{\beta} (cP_{\theta} + d) i + \frac{(\sigma_{h,k}^{(\theta)})^2 q_{\beta}^2}{c_{\theta}} \cdot \frac{1}{z} \right) \right\} \right)
 \end{aligned}
 \quad (4.2)$$

where  $\delta_{h,k}^{(\theta)} = -1$  if  $\sigma_{h,k}^{(\theta)} > 0$  and  $\delta_{h,k}^{(\theta)} = 1$  if  $\sigma_{h,k}^{(\theta)} < 0$ . The factor involving  $\delta_{h,k}^{(\theta)}$  arises because we have combined three factors into the single factor  $(c_{\theta} z / (|\sigma_{h,k}^{(\theta)}| q_{\beta}))^r$ .

In order to simplify the notation we write (4.2) as

$$\begin{aligned}
 f_{\theta} \left( \exp \left\{ 2\pi i \left( \frac{h}{k} + iz \right) \right\} \right) \\
 = \Omega_{h,k}^{(\theta)} \Psi_{h,k}^{(\theta)}(z) f_{\beta} \left( \exp \left\{ G_{h,k}^{(\theta)} i - \frac{2\pi (\sigma_{h,k}^{(\theta)})^2 q_{\beta}^2}{k c_{\beta} c_{\theta}} \cdot \frac{1}{z} \right\} \right), \quad \beta = \beta(h, k, g),
 \end{aligned}
 \quad (4.31)$$

where

$$\begin{aligned}
 \Omega_{h,k}^{(\theta)} = (\epsilon_{h,k}^{(\theta)})^{-1} \exp \left\{ \left( 1 - \delta_{h,k}^{(\theta)} \right) \frac{\pi i r}{2} \right\} \\
 \cdot \exp \left\{ - \frac{2\pi i}{k} \left( \alpha_{\theta} h + \frac{\sigma_{h,k}^{(\theta)} q_{\beta} (cP_{\theta} + d) \alpha_{\beta}}{c_{\beta}} \right) \right\},
 \end{aligned}
 \quad (4.32)$$

$$\Psi_{h,k}^{(\theta)}(z) = \left( \frac{c_{\theta} z}{|\sigma_{h,k}^{(\theta)}| q_{\beta}} \right)^r \exp \left\{ \frac{2\pi}{k} \left( \alpha_{\theta} z - \frac{\alpha_{\beta} (\sigma_{h,k}^{(\theta)})^2 q_{\beta}^2}{c_{\beta} c_{\theta}} \cdot \frac{1}{z} \right) \right\},
 \quad (4.33)$$

$$G_{h,k}^{(\theta)} = - \frac{2\pi}{k c_{\beta}} \sigma_{h,k}^{(\theta)} q_{\beta} (cP_{\theta} + d).
 \quad (4.34)$$

We note that  $|\Omega_{h,k}^{(\theta)}| = 1$  and that  $G_{h,k}^{(\theta)}$  is real.

5. A convergent series for  $a_n^{(\theta)}$ . In the following we shall let  $\sum_{h,k}^N$  designate a sum over all  $h$  and  $k$  such that  $0 \leq h < k \leq N$  and  $(h, k) = 1$ .

We let  $N$  be a positive integer and have by Cauchy's theorem

$$a_n^{(\theta)} = \frac{1}{2\pi i} \int \frac{f_{\theta}(x)}{x^{n+1}} dx$$

where we take the integral over the circle  $|x| = e^{-2\pi N^{-2}}$  in the positive direction. We make the usual Farey dissection, of order  $N$ , of the circle and set

$$x = \exp \{ -2\pi N^{-2} + 2\pi i(h/k) + 2\pi i\phi \}$$

on the arc corresponding to  $h/k$ . We then have

$$\begin{aligned} a_n^{(\vartheta)} &= \sum_{h,k}^N \int_{-\theta'}^{\theta''} \frac{f_\vartheta(\exp \{ -2\pi N^{-2} + 2\pi i(h/k) + 2\pi i\phi \})}{\exp \{ -2\pi N^{-2}n + 2\pi i(nh/k) + 2\pi i n\phi \}} d\phi \\ &= \exp \{ 2\pi N^{-2}n \} \sum_{h,k}^N \exp \left\{ -2\pi i n \frac{h}{k} \right\} \\ &\quad \cdot \int_{-\theta'}^{\theta''} f_\vartheta \left( \exp \left\{ 2\pi i \left( \frac{h}{k} + \frac{ik(N^{-2} - i\phi)}{k} \right) \right\} \right) \exp \{ -2\pi i n\phi \} d\phi \end{aligned}$$

where  $\theta'$  and  $\theta''$ , which depend on  $h$  and  $k$ , determine the end points of the Farey arcs. Equation (4.31) now yields the result

$$\begin{aligned} (5.1) \quad a_n^{(\vartheta)} &= \exp \{ 2\pi N^{-2}n \} \sum_{h,k}^N \Omega_{h,k}^{(\vartheta)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} \int_{-\theta'}^{\theta''} \Psi_{h,k}^{(\vartheta)}(k(N^{-2} - i\phi)) \\ &\quad \cdot f_\beta \left( \exp \left\{ G_{h,k}^{(\vartheta)} i - \frac{2\pi(\sigma_{h,k}^{(\vartheta)})^2 q_\beta^2}{k c_\beta c_\vartheta} \frac{1}{k(N^{-2} - i\phi)} \right\} \right) \exp \{ -2\pi i n\phi \} d\phi \end{aligned}$$

with  $\beta = \beta(h, k, g)$ .

Now  $f_\beta(x)$  is dominated by its principal part  $P_\beta(x)$  near  $x=0$ ; so we put

$$D_\beta(x) = f_\beta(x) - P_\beta(x) = \sum_{m=0}^{\infty} a_m^{(\beta)} x^m$$

and split (5.1) into two parts

$$(5.21) \quad a_n^{(\vartheta)} = Q^{(\vartheta)}(n) + R^{(\vartheta)}(n),$$

where

$$\begin{aligned} (5.22) \quad Q^{(\vartheta)}(n) &= \exp \{ 2\pi N^{-2}n \} \sum_{h,k}^N \Omega_{h,k}^{(\vartheta)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} \\ &\quad \cdot \int_{-\theta'}^{\theta''} \Psi_{h,k}^{(\vartheta)}(k(N^{-2} - i\phi)) \\ &\quad \cdot P_\beta \left( \exp \left\{ G_{h,k}^{(\vartheta)} i - \frac{2\pi(\sigma_{h,k}^{(\vartheta)})^2 q_\beta^2}{k c_\beta c_\vartheta} \frac{1}{k(N^{-2} - i\phi)} \right\} \right) \exp \{ -2\pi i n\phi \} d\phi \end{aligned}$$

and where  $R^{(\vartheta)}(n)$  is given by

$$\begin{aligned}
 R^{(\sigma)}(n) = \exp \left\{ 2\pi N^{-2}n \right\} \sum_{h,k}^N \Omega_{h,k}^{(\sigma)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} \int_{-\theta'}^{\theta''} \Psi_{h,k}^{(\sigma)}(k(N^{-2} - i\phi)) \\
 (5.23) \quad \cdot D_{\beta} \left( \exp \left\{ G_{h,k}^{(\sigma)} i - \frac{2\pi(\sigma_{h,k}^{(\sigma)})^2 q_{\beta}^2}{k c_{\beta} c_{\sigma}} \frac{1}{k(N^{-2} - i\phi)} \right\} \right) \\
 \cdot \exp \{ -2\pi i n \phi \} d\phi.
 \end{aligned}$$

We can now make an estimate of  $R^{(\sigma)}(n)$ . From the theory of Farey fractions we have

$$(5.3) \quad \frac{1}{2kN} < \theta' < \frac{1}{kN}, \quad \frac{1}{2kN} < \theta'' < \frac{1}{kN};$$

hence for  $-\theta' \leq \phi \leq \theta''$  we have

$$\begin{aligned}
 R(k(N^{-2} - i\phi)) &= kN^{-2}, \quad R\left(\frac{1}{k(N^{-2} - i\phi)}\right) \geq \frac{k}{2}, \\
 |k(N^{-2} - i\phi)| &\leq 2^{1/2}N^{-1}.
 \end{aligned}$$

Using these results we find, by (4.33),

$$\begin{aligned}
 &\left| \Psi_{h,k}^{(\sigma)}(k(N^{-2} - i\phi)) D_{\beta} \left( \exp \left\{ G_{h,k}^{(\sigma)} i - \frac{2\pi(\sigma_{h,k}^{(\sigma)})^2 q_{\beta}^2}{k c_{\beta} c_{\sigma}} \frac{1}{k(N^{-2} - i\phi)} \right\} \right) \right| \\
 &\leq \frac{c_{\sigma}^r}{|\sigma_{h,k}^{(\sigma)}|^r q_{\beta}^r} 2^{r/2} N^{-r} \exp \left\{ 2\pi \alpha_{\sigma} N^{-2} - \frac{\pi \alpha_{\beta} (\sigma_{h,k}^{(\sigma)})^2 q_{\beta}^2}{c_{\beta} c_{\sigma}} \right\} \\
 &\quad \cdot \sum_{m=0}^{\infty} |a_m^{(\beta)}| \exp \left\{ -\frac{\pi (\sigma_{h,k}^{(\sigma)})^2 q_{\beta}^2}{c_{\beta} c_{\sigma}} m \right\} \\
 &\leq \frac{c_{\sigma}^r}{q_{\beta}^r} 2^{r/2} N^{-r} \exp \{ 2\pi \alpha_{\sigma} N^{-2} \} \sum_{m=0}^{\infty} |a_m^{(\beta)}| \exp \left\{ -\frac{\pi q_{\beta}^2}{c_{\beta} c_{\sigma}} m \right\} \\
 &\leq KN^{-r} \exp \{ 2\pi \alpha_{\sigma} N^{-2} \}
 \end{aligned}$$

where we have used (4.14) and the fact that the series for  $D_{\beta}(x)$  converges for  $|x| < 1$ . Combining this result with (5.23), we have

$$\begin{aligned}
 |R^{(\sigma)}(n)| &\leq \exp \{ 2\pi N^{-2}n \} KN^{-r} \exp \{ 2\pi \alpha_{\sigma} N^{-2} \} \\
 (5.4) \quad &\cdot \sum_{h,k}^N \int_{-\theta'}^{\theta''} d\phi = KN^{-r} \exp \{ 2\pi(n + \alpha_{\sigma})N^{-2} \}.
 \end{aligned}$$

The determination of  $a_n^{(\sigma)}$  now rests entirely on the evaluation of  $Q^{(\sigma)}(n)$ . To accomplish this we set  $w = N^{-2} - i\phi$  in (5.22) and have, by (3.13), the relation

$$\begin{aligned}
 Q^{(\sigma)}(n) &= \exp \{2\pi N^{-2}n\} \sum_{h,k}^N \Omega_{h,k}^{(\sigma)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} \\
 &\quad \cdot \int_{N^{-2}+i\theta'}^{N^{-2}-i\theta'''} i \Psi_{h,k}^{(\sigma)}(kw) \sum_{\nu=1}^{\mu_\beta} a_{-\nu}^{(\beta)} \\
 (5.51) \quad &\cdot \exp \left\{ -\nu G_{h,k}^{(\sigma)} i + \frac{2\pi(\sigma_{h,k}^{(\sigma)})^2 q_\beta^2}{k^2 c_\beta c_\sigma w} \frac{\nu}{kw} \right\} \exp \{2\pi n w - 2\pi n N^{-2}\} dw \\
 &= \sum_{h,k}^N \Omega_{h,k}^{(\sigma)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} \sum_{\nu=1}^{\mu_\beta} a_{-\nu}^{(\beta)} \exp \{ -\nu G_{h,k}^{(\sigma)} i \} I_{h,k,\nu}^{(\sigma)}(n),
 \end{aligned}$$

where

$$(5.52) \quad I_{h,k,\nu}^{(\sigma)}(n) = \frac{1}{i} \int_{N^{-2}-i\theta'''}^{N^{-2}+i\theta'} \Psi_{h,k}^{(\sigma)}(kw) \exp \left\{ \frac{2\pi\nu(\sigma_{h,k}^{(\sigma)})^2 q_\beta^2}{k^2 c_\beta c_\sigma w} + 2\pi n w \right\} dw$$

and where the inner sum is to be taken as zero if  $\mu_\beta = 0$ . By (4.33) we may write (5.52) as

$$(5.53) \quad I_{h,k,\nu}^{(\sigma)}(n) = k^r \left( \frac{c_\sigma}{|\sigma_{h,k}^{(\sigma)}| q_\beta} \right)^r \frac{1}{i} \int_{N^{-2}-i\theta'''}^{N^{-2}+i\theta'} w^r \exp \left\{ 2\pi(\alpha_\sigma + n)w + \frac{2\pi(\sigma_{h,k}^{(\sigma)})^2 q_\beta^2}{k^2 c_\beta c_\sigma w} (\nu - \alpha_\beta) \right\}.$$

We restrict our considerations to those  $a_n^{(\sigma)}$  for which  $\alpha_\sigma + n > 0$ . That is, we leave undetermined all  $a_0^{(\sigma)}$  for which  $\alpha_\sigma = 0$  but determine all other  $a_n^{(\sigma)}$ . This restriction is necessary in order that a certain integral shall converge. We cut the  $w$ -plane from 0 to  $-\infty$  along the negative real axis and consider a path of integration encircling the cut in the positive sense and connecting the points

$$-\infty, -\epsilon, -\epsilon - i\theta''', N^{-2} - i\theta''', N^{-2} + i\theta', -\epsilon + i\theta', -\epsilon, -\infty$$

by straight lines where we take  $0 < \epsilon < N^{-2}$ . Then we can write

$$\begin{aligned}
 k^{-r} \left( \frac{c_\sigma}{|\sigma_{h,k}^{(\sigma)}| q_\beta} \right)^{-r} I_{h,k,\nu}^{(\sigma)}(n) &= \frac{1}{i} \int_{-\infty}^{(0^+)} - \frac{1}{i} \int_{-\infty}^{-\epsilon} - \frac{1}{i} \int_{-\epsilon}^{-\epsilon - i\theta'''} \\
 &\quad - \frac{1}{i} \int_{-\epsilon - i\theta'''}^{N^{-2} - i\theta'''} - \frac{1}{i} \int_{N^{-2} - i\theta'''}^{-\epsilon + i\theta'} \\
 (5.61) \quad &\quad - \frac{1}{i} \int_{-\epsilon + i\theta'}^{-\infty} - \frac{1}{i} \int_{-\infty}^{-\epsilon} \\
 &= L_{h,k,\nu}^{(\sigma)}(n) - J_1 - J_2 - J_3 - J_4 - J_5 - J_6,
 \end{aligned}$$

say. All these integrals have the same integrand

$$w^r \exp \left\{ 2\pi(\alpha_\theta + n)w + \frac{2\pi(\sigma_{h,k}^{(\theta)})^2 q \beta^2}{k^2 c_\beta c_\theta w} (\nu - \alpha_\beta) \right\}.$$

The integral  $J_1$  is to be taken on the border below the cut,  $J_6$  above the cut. In the integral  $J_2$  we have  $w = -\epsilon + iv$ ,  $0 \leq v \leq -\theta''$ , and

$$R(w) = -\epsilon, \quad R(1/w) = -\epsilon/(\epsilon^2 + v^2) < 0, \quad |w| \leq 2^{1/2} k^{-1} N^{-1},$$

and therefore

$$(5.62) \quad |J_2| \leq \theta'' 2^{r/2} k^{-r} N^{-r} < 2^{r/2} k^{-r-1} N^{-r-1}.$$

Similarly we have

$$(5.63) \quad |J_5| \leq 2^{r/2} k^{-r-1} N^{-r-1}.$$

In the integral  $J_3$  we have  $w = u - i\theta''$ ,  $-N^{-2} < -\epsilon \leq u \leq N^{-2}$ , and

$$R(w) = u \leq N^{-2}, \quad R(1/w) \leq 4k^2, \quad |w| \leq 2^{1/2} k^{-1} N^{-1}$$

and therefore, using (4.14),

$$\begin{aligned} |J_3| &\leq (N^{-2} + \epsilon) 2^{r/2} k^{-r} N^{-r} \exp \left\{ 2\pi(\alpha_\theta + n)N^{-2} + \frac{8\pi(\sigma_{h,k}^{(\theta)})^2 q \beta^2}{c_\theta c_\beta} (\nu - \alpha_\beta) \right\} \\ (5.64) \quad &\leq 2N^{-2} 2^{r/2} k^{-r} N^{-r} \exp \left\{ 2\pi(\alpha_\theta + n)N^{-2} + \frac{8\pi c_\theta q \beta^2 (\mu_\theta - \alpha_\beta)}{c_\beta} \right\} \\ &\leq K k^{-r-1} N^{-r-1} \exp \{ 2\pi(\alpha_\theta + n)N^{-2} \} \end{aligned}$$

and similarly

$$(5.65) \quad |J_4| \leq K k^{-r-1} N^{-r-1} \exp \{ 2\pi(\alpha_\theta + n)N^{-2} \}.$$

Finally we have

$$\begin{aligned} J_1 + J_6 &= \frac{1}{i} \int_{-\infty}^{-\epsilon} |w|^r \exp \{ -\pi i r \} \\ &\quad \cdot \exp \left\{ 2\pi(\alpha_\theta + n)w + \frac{2\pi(\sigma_{h,k}^{(\theta)})^2 q \beta^2 (\nu - \alpha_\beta)}{k^2 c_\beta c_\theta w} \right\} dw \\ (5.66) \quad &+ \frac{1}{i} \int_{-\epsilon}^{-\infty} |w|^r \exp \{ \pi i r \} \\ &\quad \cdot \exp \left\{ 2\pi(\alpha_\theta + n)w + \frac{2\pi(\sigma_{h,k}^{(\theta)})^2 q \beta^2 (\nu - \alpha_\beta)}{k^2 c_\beta c_\theta w} \right\} dw \\ &= -2 \sin \pi r \int_{\epsilon}^{\infty} t^r \exp \left\{ -2\pi(\alpha_\theta + n)t - \frac{2\pi(\sigma_{h,k}^{(\theta)})^2 q \beta^2 (\nu - \alpha_\beta)}{k^2 c_\beta c_\theta t} \right\} dt \end{aligned}$$

where our restriction  $\alpha_\theta + n > 0$  ensures the convergence of the integrals. Combining (5.61), (5.62), (5.63), (5.64), (5.65), (5.66), and letting  $\epsilon \rightarrow 0$ , we have

$$I_{h,k,r}^{(\theta)}(n) = k^r \left( \frac{c_\theta}{|\sigma_{h,k}^{(\theta)}| q_\beta} \right)^r L_{h,k,r}^{(\theta)}(n) + 2k^r \left( \frac{c_\theta}{|\sigma_{h,k}^{(\theta)}| q_\beta} \right)^r \cdot \sin \pi r \int_0^\infty t^r \exp \left\{ -2\pi(\alpha_\theta + n)t - \frac{2\pi(\sigma_{h,k}^{(\theta)})^2 q_\beta^2 (\nu - \alpha_\beta)}{k^2 c_\beta c_\theta t} \right\} dt \\ + O \left( k^r \left( \frac{c_\theta}{|\sigma_{h,k}^{(\theta)}| q_\beta} \right)^r k^{-r-1} N^{-r-1} \exp \{ 2\pi(\alpha_\theta + n)N^{-2} \} \right).$$

Introducing this into (5.51) and using (4.14), we obtain

$$Q^{(\theta)}(n) = \sum_{h,k}^N \Omega_{h,k}^{(\theta)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} \sum_{\nu=1}^{\mu_\beta} a_{-\nu}^{(\beta)} \exp \{ -\nu G_{h,k}^{(\theta)} i \} \\ \cdot k^r \left( \frac{c_\theta}{|\sigma_{h,k}^{(\theta)}| q_\beta} \right)^r \{ L_{h,k,r}^{(\theta)}(n) + M_{h,k,r}^{(\theta)}(n) \} \\ + O \left( N^{-r-1} \exp \{ 2\pi(\alpha_\theta + n)N^{-2} \} \sum_{\nu=1}^{\mu_\beta} |\alpha_{-\nu}^{(\beta)}| \sum_{h,k}^N k^{-1} \right) \quad (5.71)$$

with

$$L_{h,k,r}^{(\theta)}(n) = \frac{1}{i} \int_{-\infty}^{(0^+)} w^r \exp \left\{ 2\pi(\alpha_\theta + n)w + \frac{2\pi(\sigma_{h,k}^{(\theta)})^2 q_\beta^2 (\nu - \alpha_\beta)}{k^2 c_\beta c_\theta w} \right\} dw \quad (5.72)$$

and

$$M_{h,k,r}^{(\theta)}(n) = 2 \sin \pi r \int_0^\infty t^r \exp \left\{ -2\pi(\alpha_\theta + n)t - \frac{2\pi(\sigma_{h,k}^{(\theta)})^2 q_\beta^2 (\nu - \alpha_\beta)}{k^2 c_\beta c_\theta t} \right\} dt. \quad (5.73)$$

Now since  $\sum_{h,k}^N k^{-1} \leq N$ , we have, by (5.71), (5.21), and (5.4),

$$a_n^{(\theta)} = \sum_{h,k}^N \Omega_{h,k}^{(\theta)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} k^r \left( \frac{c_\theta}{|\sigma_{h,k}^{(\theta)}| q_\beta} \right)^r \sum_{\nu=1}^{\mu_\beta} a_{-\nu}^{(\beta)} \cdot \exp \{ -\nu G_{h,k}^{(\theta)} i \} \{ L_{h,k,r}^{(\theta)}(n) + M_{h,k,r}^{(\theta)}(n) \} \\ + O(N^{-r} \exp \{ 2\pi(\alpha_\theta + n)N^{-2} \}).$$

If we keep  $n$  fixed and let  $N$  tend to infinity, the error term tends to zero, since  $r > 0$ . Hence the series thus obtained converges and we have

$$a_n^{(\theta)} = \sum_{h,k}^\infty \Omega_{h,k}^{(\theta)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} k^r \left( \frac{c_\theta}{|\sigma_{h,k}^{(\theta)}| q_\beta} \right)^r \cdot \sum_{\nu=1}^{\mu_\beta} a_{-\nu}^{(\beta)} \exp \{ -\nu G_{h,k}^{(\theta)} i \} \{ L_{h,k,r}^{(\theta)}(n) + M_{h,k,r}^{(\theta)}(n) \}. \quad (5.8)$$

6. **Evaluation of the integrals.** We now express  $L_{h,k,v}^{(\theta)}(n)$  and  $M_{h,k,v}^{(\theta)}(n)$  in terms of Bessel functions. In (5.72) the change of variable  $u = 2\pi(\alpha_\theta + n)w$  yields

$$L_{h,k,v}^{(\theta)}(n) = \frac{1}{i(2\pi(\alpha_\theta + n))^{r+1}} \cdot \int_{-\infty}^{(0^+)} u^r \exp \left\{ u + \frac{4\pi^2(\sigma_{h,k}^{(\theta)})^2 q_\beta^2 (\nu - \alpha_\beta)(\alpha_\theta + n)}{k^2 c_\beta c_\theta u} \right\} du;$$

hence we have, by a well known formula,\*

$$(6.1) \quad L_{h,k,v}^{(\theta)}(n) = \frac{2\pi |\sigma_{h,k}^{(\theta)}|^{r+1} q_\beta^{r+1} (\nu - \alpha_\beta)^{(r+1)/2}}{k^{r+1} c_\beta^{(r+1)/2} c_\theta^{(r+1)/2} (\alpha_\theta + n)^{(r+1)/2}} \cdot I_{-r-1} \left( \frac{4\pi |\sigma_{h,k}^{(\theta)}| q_\beta ((\nu - \alpha_\beta)(\alpha_\theta + n))^{1/2}}{k(c_\beta c_\theta)^{1/2}} \right),$$

where  $I_m(z)$  is the Bessel function of the first kind with purely imaginary argument.

In (5.73) we set  $v = 2\pi(\alpha_\theta + n)t$  and have

$$(6.2) \quad M_{h,k,v}^{(\theta)}(n) = \frac{2 \sin \pi r}{(2\pi(\alpha_\theta + n))^{r+1}} \cdot \int_0^\infty v^r \exp \left\{ -v - \frac{4\pi^2(\sigma_{h,k}^{(\theta)})^2 q_\beta^2 (\nu - \alpha_\beta)(\alpha_\theta + n)}{k^2 c_\beta c_\theta v} \right\} dv \\ = \frac{4 |\sigma_{h,k}^{(\theta)}|^{r+1} q_\beta^{r+1} (\nu - \alpha_\beta)^{(r+1)/2} \sin \pi r}{k^{r+1} c_\beta^{(r+1)/2} c_\theta^{(r+1)/2} (\alpha_\theta + n)^{(r+1)/2}} \cdot K_{-r-1} \left( \frac{4\pi |\sigma_{h,k}^{(\theta)}| q_\beta ((\nu - \alpha_\beta)(\alpha_\theta + n))^{1/2}}{k(c_\beta c_\theta)^{1/2}} \right),$$

where  $K_m(z)$  is the Bessel function of the third kind with purely imaginary argument.† Now adding (6.1) and (6.2) and simplifying by means of the formula‡

$$\sin \pi \nu K_\nu(z) + \frac{\pi}{2} I_\nu(z) = \frac{\pi}{2} I_{-\nu}(z),$$

we find the relation

\* G. N. Watson, *Theory of Bessel Functions*, Cambridge, 1922, p. 181, (1).

† Watson, loc. cit., p. 183, (15).

‡ Watson, loc. cit., p. 78, (6).

$$\begin{aligned}
 k^r \left( \frac{c_g}{|\sigma_{h,k}^{(g)}| q_g} \right)^r \{ L_{h,k,v}^{(g)}(n) + M_{h,k,v}^{(g)}(n) \} \\
 = \frac{2\pi |\sigma_{h,k}^{(g)}| q_g (v - \alpha_g)^{(r+1)/2} c_g^{(r-1)/2}}{k c_g^{(r+1)/2} (\alpha_g + n)^{(r+1)/2}} \\
 \cdot I_{r+1} \left( \frac{4\pi |\sigma_{h,k}^{(g)}| q_g ((v - \alpha_g)(\alpha_g + n))^{1/2}}{k (c_g c_g)^{1/2}} \right).
 \end{aligned}$$

Using this result in (5.8) we have the following theorem:

**THEOREM 1.** Let  $F(\tau)$  be a modular form of positive dimension  $r$  belonging to the subgroup  $\gamma$ , and let its expansions (3.11) have only a finite number of terms with negative exponent. Then the coefficients  $a_n^{(g)}$  of (3.12) for which  $\alpha_g + n > 0$  are given by

$$\begin{aligned}
 a_n^{(g)} = & \frac{2\pi c_g^{(r-1)/2}}{(\alpha_g + n)^{(r+1)/2}} \sum_{h,k} \Omega_{h,k}^{(g)} \exp \left\{ -2\pi i n \frac{h}{k} \right\} \\
 (6.3) \quad & \cdot \sum_{s=1}^{\mu_g} a_{-s}^{(g)} \exp \left\{ -s G_{h,k}^{(g)} i \right\} \frac{|\sigma_{h,k}^{(g)}| q_g (v - \alpha_g)^{(r+1)/2}}{k c_g^{(r+1)/2}} \\
 & \cdot I_{r+1} \left( \frac{4\pi |\sigma_{h,k}^{(g)}| q_g ((v - \alpha_g)(\alpha_g + n))^{1/2}}{k (c_g c_g)^{1/2}} \right)
 \end{aligned}$$

where  $\beta = \beta(h, k, g)$  is defined in §4;  $c_\beta$  and  $c_g$  are determined by (3.11);  $\alpha_\beta$  and  $\alpha_g$  by (3.12);  $\Omega_{h,k}^{(g)}$  is defined in (4.32); and  $G_{h,k}^{(g)}$  in (4.34). The  $F^*(\tau)$  of (1.32) are then given by (3.23).

**7. Application to a particular subgroup.** Many of the constants in (6.3) depend merely on the subgroup  $\gamma$  and are otherwise independent of the choice of  $F(\tau)$ . For this reason we can simplify the expression if we choose some particular subgroup  $\gamma$ . As an example we consider the subgroup consisting of all transformations

$$(7.1) \quad \tau' = \frac{a\tau + b}{c\tau + d}, \quad a \equiv d \equiv 1, \quad c \equiv 0 \pmod{2}.$$

For the region  $R$  we may take that part of the plane which is above the two circles  $|\tau - 1/4| = 1/4$  and  $|\tau - 3/4| = 1/4$  and which is within the circle  $|\tau - 1/2| = 1/2$ . Then  $R$  has the three parabolic points 0,  $1/2$ , and 1, but 0 and 1 are congruent under  $\gamma$ ; so we have  $s=2$  and take  $P_1=0$ ,  $P_2=1/2$ . Then we have

$$p_1 = 0, \quad q_1 = 1, \quad p_2 = 1, \quad q_2 = 2.$$

Corresponding to the transformations (3.11) we have the two transformations

$$\tau' = \frac{\tau}{2\tau + 1}, \quad \tau' = \frac{3\tau - 1}{4\tau - 1}$$

which may be written in the form

$$\frac{1}{\tau'} = \frac{1}{\tau} + 2, \quad \frac{1}{\tau' - 1/2} = \frac{1}{\tau - 1/2} + 4;$$

so we have  $c_1 = 2$  and  $c_2 = 4$ . The expansions (3.12) can now be written as

$$(7.21) \quad F(\tau) = (-i\tau)^r x^{\alpha_1} \sum_{m=-\mu_1}^{\infty} a_m^{(1)} x^m, \quad x = \exp \left\{ -\frac{\pi i}{\tau} \right\};$$

$$(7.22) \quad F(\tau) = (-i(\tau - 1/2))^r x^{\alpha_2} \sum_{m=-\mu_2}^{\infty} a_m^{(2)} x^m, \quad x = \exp \left\{ -\frac{\pi i}{2\tau - 1} \right\}.$$

In order to determine  $\beta(h, k, g)$  we find transformations (7.1) which transform  $P_g - k/c_g h$  into  $P_\beta$ . We define  $h'$ ,  $h''$ ,  $h'''$ , and  $h^{iv}$  as any solutions of

$$(7.31) \quad kh' \equiv 1 \pmod{2h}, \quad h' > 0, \text{ for } k \equiv 1 \pmod{2};$$

$$(7.32) \quad kh'' \equiv -1 \pmod{h}, \quad h'' > 0, \text{ for } k \equiv 0 \pmod{2};$$

$$(7.33) \quad (k - 2h)h''' \equiv 1 \pmod{4h}, \quad h''' > 0, \text{ for } k \equiv 1 \pmod{2};$$

$$(7.34) \quad \left( \frac{k}{2} - h \right) h^{iv} \equiv 1 \pmod{2h}, \quad h^{iv} > 0, \text{ for } k \equiv 0 \pmod{4};$$

and may take for  $a$ ,  $b$ ,  $c$ , and  $d$  the values given in the following tables.

$g$	$k$	$a$	$b$	$c$
1	$\equiv 1 \pmod{2}$	$h'$	$\frac{h'k - 1}{2h}$	$2h' + 2h$
1	$\equiv 0 \pmod{2}$	$h$	$\frac{k}{2}$	$2h''$
2	$\equiv 1 \pmod{2}$	$h'''$	$\frac{(k - 2h)h''' - 1}{4h}$	$2h''' + 4h$
2	$\equiv 0 \pmod{4}$	$h^{iv}$	$\frac{\left( \frac{k}{2} - h \right) h^{iv} - 1}{2h}$	$2h^{iv} + 2h$
2	$\equiv 2 \pmod{4}$	$h$	$\frac{k - 2h}{4}$	$4h''$

$d$	$\beta$	$\sigma_{h,k}^{(\sigma)}$	$\delta_{h,k}^{(\sigma)}$	$G_{h,k}^{(\sigma)}$
$\frac{h'k-1}{h} + k$	2	-1	1	$\left(\frac{h'k-1}{kh} + 1\right)\pi$
$\frac{1+h''k}{h}$	1	2	-1	$-\left(\frac{2+2h''k}{kh}\right)\pi$
$\frac{(k-2h)h'''-1}{2h} - 2h + k$	2	-1	1	$\left(\frac{h'''k-1}{2kh} + 1\right)\pi$
$\frac{\left(\frac{k}{2}-h\right)h^{iv}-1}{h} + \frac{k}{2} - h$	2	-2	1	$\left(\frac{h^{iv}k-2}{kh} + 1\right)\pi$
$\frac{kh''+1}{h} - 2h''$	1	4	-1	$-\left(\frac{4h''k+4}{kh}\right)\pi$

The values of  $\sigma_{h,k}^{(\sigma)}$  were found by means of (4.13) and  $G_{h,k}^{(\sigma)}$  from (4.34).

The series (6.3) can be shown to be absolutely convergent; so we may rearrange the terms. Doing this and making use of (4.32) and the tables, we get

$$\begin{aligned}
 a_n^{(1)} &= 2\pi \sum_{\nu=1}^{\mu_1} a_{-\nu}^{(1)} \sum_{\substack{k \equiv 0 \pmod{2} \\ k > 0}} \frac{1}{k} A_{k,\nu}^{(1)}(n) \\
 &\cdot \left(\frac{\nu - \alpha_1}{n + \alpha_1}\right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_1)(n + \alpha_1))^{1/2}}{k} \right) \\
 (7.41) \quad &+ \frac{\pi}{2^{(r-1)/2}} \sum_{\nu=1}^{\mu_2} a_{-\nu}^{(2)} \sum_{\substack{k \equiv 1 \pmod{2} \\ k > 0}} \frac{1}{k} A_{k,\nu}^{(1)}(n) \\
 &\cdot \left(\frac{\nu - \alpha_2}{n + \alpha_1}\right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_2)(n + \alpha_1))^{1/2}}{k2^{1/2}} \right),
 \end{aligned}$$

where

$$\begin{aligned}
 A_{k,\nu}^{(1)}(n) &= \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \epsilon \left( h, \frac{k}{2}, 2h'', \frac{1+h''k}{h} \right)^{-1} \exp \{ \pi i r \} \\
 (7.42) \quad &\cdot \exp \left\{ -\frac{2\pi i}{k} \alpha_1 h - (\alpha_1 - \nu) \left( \frac{2h''k+2}{kh} \right) \pi i - 2\pi i n \frac{h}{k} \right\}, \\
 &\text{if } k \equiv 0 \pmod{2},
 \end{aligned}$$

$$\begin{aligned}
 A_{k,v}^{(1)}(n) &= \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \epsilon \left( h', \frac{h'k-1}{2}, 2h'+2h, \frac{h'k-1}{h} + k \right)^{-1} \\
 (7.43) \quad &\cdot \exp \left\{ -\frac{2\pi i}{k} \alpha_1 h + (\alpha_2 - \nu) \left( \frac{h'k-1}{kh} + 1 \right) \pi i - 2\pi i n \frac{h}{k} \right\}, \\
 &\text{if } k \equiv 1 \pmod{2},
 \end{aligned}$$

if  $g=1$  for all  $n$  such that  $\alpha_1+n>0$ . Similarly, if  $g=2$  and  $\alpha_2+n>0$ , we have

$$\begin{aligned}
 a_n^{(2)} &= 2^{(r+3)/2} \pi \sum_{\nu=1}^{\mu_1} a_{-\nu}^{(1)} \sum_{\substack{k \equiv 2 \pmod{4} \\ k>0}} \frac{1}{k} A_{k,v}^{(2)}(n) \left( \frac{\nu - \alpha_1}{n + \alpha_2} \right)^{(r+1)/2} \\
 (7.44) \quad &\cdot I_{r+1} \left( \frac{8\pi((\nu - \alpha_1)(n + \alpha_2))^{1/2}}{k2^{1/2}} \right) + 2\pi \sum_{\nu=1}^{\mu_2} a_{-\nu}^{(2)} \sum_{\substack{k \equiv 0 \pmod{4} \\ k>0}} \frac{1}{k} A_{k,v}^{(2)}(n) \\
 &\cdot \left( \frac{\nu - \alpha_2}{n + \alpha_2} \right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_2)(n + \alpha_2))^{1/2}}{k} \right) \\
 &+ \pi \sum_{\nu=1}^{\mu_3} a_{-\nu}^{(2)} \sum_{\substack{k \equiv 1 \pmod{2} \\ k>0}} \frac{1}{k} A_{k,v}^{(2)}(n) \\
 &\cdot \left( \frac{\nu - \alpha_2}{n + \alpha_2} \right)^{(r+1)/2} I_{r+1} \left( \frac{2\pi((\nu - \alpha_2)(n + \alpha_2))^{1/2}}{k} \right),
 \end{aligned}$$

where

$$\begin{aligned}
 A_{k,v}^{(2)}(n) &= \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \epsilon \left( h, \frac{k-2h}{4}, 4h'', \frac{kh''+1}{h} - 2h'' \right)^{-1} \exp \{ \pi i r \} \\
 (7.45) \quad &\cdot \exp \left\{ -\frac{2\pi i}{k} \alpha_2 h - (\alpha_1 - \nu) \left( \frac{2h''k+2}{kh} \right) \pi i - 2\pi i n \frac{h}{k} \right\}, \\
 &\text{if } k \equiv 2 \pmod{4},
 \end{aligned}$$

$$\begin{aligned}
 A_{k,v}^{(2)}(n) &= \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \epsilon \left( h^{iv}, \frac{(k/2-h)h^{iv}-1}{2h}, 2h^{iv}+2h, \frac{(k/2-h)h^{iv}-1}{h} + \frac{k}{2} - h \right)^{-1} \\
 (7.46) \quad &\cdot \exp \left\{ -\frac{2\pi i}{k} \alpha_2 h + (\alpha_2 - \nu) \left( \frac{h^{iv}k-2}{kh} + 1 \right) \pi i - 2\pi i n \frac{h}{k} \right\}, \\
 &\text{if } k \equiv 0 \pmod{4},
 \end{aligned}$$

$$\begin{aligned}
 & A_{k,v}^{(2)}(n) \\
 &= \sum_{\substack{0 \leq k < k \\ (h,k)=1}} \epsilon \left( h''', \frac{(k-2h)h'''-1}{4h}, 2h''' + 4h, \frac{(k-2h)h'''-1}{2h} - 2h + k \right)^{-1} \\
 & \cdot \exp \left\{ -\frac{2\pi i}{k} \alpha_2 h + (\alpha_2 - \nu) \left( \frac{h'''k-1}{2kh} + 1 \right) \pi i - 2\pi i n \frac{h}{k} \right\}, \\
 & \text{if } k \equiv 1 \pmod{2}.
 \end{aligned}
 \tag{7.47}$$

Equation (7.44) can be given a more symmetric form by changing the indices of summation. In the first sum we set  $k=2l$  in the second  $k=l$ , and in the third  $k=l/2$ . Then we find

$$\begin{aligned}
 a_n^{(2)} &= 2^{(r+1)/2} \pi \sum_{\nu=1}^{\mu_1} a_{-\nu}^{(1)} \sum_{\substack{l \equiv 1 \pmod{2} \\ l > 0}} \frac{1}{l} A_{2l,\nu}^{(2)}(n) \\
 & \cdot \left( \frac{\nu - \alpha_1}{n + \alpha_2} \right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_1)(n + \alpha_2))^{1/2}}{l^{1/2}} \right) \\
 & + 2\pi \sum_{\nu=1}^{\mu_2} a_{-\nu}^{(2)} \sum_{\substack{l \equiv 0 \pmod{4} \\ l > 0}} \frac{1}{l} A_{l,\nu}^{(2)}(n) \\
 & \cdot \left( \frac{\nu - \alpha_2}{n + \alpha_2} \right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_2)(n + \alpha_2))^{1/2}}{l} \right) \\
 & + 2\pi \sum_{\nu=1}^{\mu_2} a_{-\nu}^{(2)} \sum_{\substack{l \equiv 2 \pmod{4} \\ l > 0}} \frac{1}{l} A_{l/2,\nu}^{(2)}(n) \\
 & \cdot \left( \frac{\nu - \alpha_2}{n + \alpha_2} \right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_2)(n + \alpha_2))^{1/2}}{l} \right) \\
 & = 2^{(r+1)/2} \pi \sum_{\nu=1}^{\mu_1} a_{-\nu}^{(1)} \sum_{\substack{l \equiv 1 \pmod{2} \\ l > 0}} \frac{1}{l} A_{l',\nu}^{(2)}(n) \\
 & \cdot \left( \frac{\nu - \alpha_1}{n + \alpha_2} \right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_1)(n + \alpha_2))^{1/2}}{l^{1/2}} \right) \\
 & + 2\pi \sum_{\nu=1}^{\mu_2} a_{-\nu}^{(2)} \sum_{\substack{l \equiv 0 \pmod{2} \\ l > 0}} \frac{1}{l} A_{l',\nu}^{(2)}(n) \\
 & \cdot \left( \frac{\nu - \alpha_2}{n + \alpha_2} \right)^{(r+1)/2} I_{r+1} \left( \frac{4\pi((\nu - \alpha_2)(n + \alpha_2))^{1/2}}{l} \right)
 \end{aligned}
 \tag{7.48}$$

where

$$(7.49) \quad l' = 2l \text{ if } l \equiv 1 \pmod{2}, \quad l' = l \text{ if } l \equiv 0 \pmod{4}, \quad l' = l/2 \text{ if } l \equiv 2 \pmod{4}.$$

We now find the functions  $F^*(\tau)$  of (1.32). If  $C$  is odd and  $D$  even, we take

$$\tau' = \frac{\pm B\tau \mp A}{\pm D\tau \mp C}$$

for the transformation (3.21) where we take the upper or lower signs to make  $\pm D \geq 0$ . This transformation takes  $P_1=0$  into  $A/C$  and it belongs to  $\gamma$ . From (3.22) we find  $\kappa = \mp 1$  and from (3.23) and (3.24) we have

$$(7.51) \quad F^*(\tau) = e^{\pi i \alpha_1 \tau} f_1(e^{\pi i \tau}) = e^{\pi i \alpha_1 \tau} \sum_{m=-\mu_1}^{\infty} a_m^{(1)} e^{\pi i m \tau},$$

$$\epsilon^* = \epsilon(\pm B, \mp A, \pm D, \mp C) e^{\mp \pi i r/2}.$$

Similarly if  $C$  and  $D$  are both odd, we use the transformation

$$\tau' = \frac{\pm (B+A)\tau \mp A}{\pm (D+C)\tau \mp C}, \quad \pm (D+C) \geq 0,$$

and obtain

$$(7.52) \quad F^*(\tau) = e^{\pi i \alpha_1 \tau} f_1(e^{-\pi i \tau}) = e^{\pi i \alpha_1 \tau} \sum_{m=-\mu_1}^{\infty} (-1)^m a_m^{(1)} e^{\pi i m \tau},$$

$$\epsilon^* = \epsilon(\pm (B+A), \mp A, \pm (D+C), \mp C) e^{\mp \pi i r/2} e^{-\pi i \alpha_1}.$$

Finally if  $C$  is even, we use the transformation

$$\tau' = \frac{\pm (A+2B)\tau \mp (A+B)}{\pm (C+2D)\tau \mp (C+D)}, \quad \pm (C+2D) \geq 0,$$

which takes  $P_2=1/2$  into  $A/C$  and find

$$(7.53) \quad F^*(\tau) = 2^{-r} e^{2\pi i \alpha_2 \tau} f_2(e^{-\pi i \tau}) = 2^{-r} e^{2\pi i \alpha_2 \tau} \sum_{m=-\mu_2}^{\infty} (-1)^m a_m^{(2)} e^{2\pi i m \tau},$$

$$\epsilon^* = \epsilon(\pm (A+2B), \mp (A+B), \pm (C+2D), \mp (C+D)) e^{\mp \pi i r/2} e^{-\pi i \alpha_2}.$$

Making use of (1.32) we summarize the results of this section in the following theorem:

**THEOREM 2.** *If  $F(\tau)$  is a modular form of positive dimension belonging to the subgroup (7.1) and if its expansions (7.21) and (7.22) have only a finite number of terms with negative exponent, then the values of  $a_n^{(g)}$  for  $\alpha_0 + n > 0$ , ( $g=1, 2$ ), are determined as functions of  $a_n^{(g)}$  with  $n < 0$  by (7.41) and (7.48).*

Also we have

$$(7.6) \quad F\left(\frac{A\tau + B}{C\tau + D}\right) = \epsilon^*(-i(C\tau + D))^{-r} F^*(\tau)$$

for any modular transformation  $\tau' = (A\tau + B)/(C\tau + D)$  where  $\epsilon^*$  and  $F^*(\tau)$  are given by (7.51) for  $C$  odd,  $D$  even; by (7.52) for  $C$  odd,  $D$  odd; and by (7.53) for  $C$  even.

8. **An example.** As a particular example we consider the function

$$(8.11) \quad F(\tau) = (\vartheta_2(0|\tau))^{-1}$$

which belongs to the subgroup discussed in §7. From the theory of the theta-functions we take only the following results:†

$$(8.12) \quad \vartheta_2\left(0 \left| \frac{a\tau + b}{c\tau + d} \right. \right) = \epsilon'(\pm c\tau \pm d)^{1/2} \vartheta_2(0|\tau)$$

if  $a$  and  $d$  are odd and  $c$  is even where we take the upper or lower sign to make  $\pm d > 0$  and where

$$(8.13) \quad \epsilon' = \left( \frac{\pm c}{\pm d} \right) \exp \left\{ \frac{\pi i}{4} (bd \pm 5d - 5) \right\};$$

$$(8.14) \quad \vartheta_2\left(0 \left| -\frac{1}{\tau} \right. \right) = (-i\tau)^{1/2} \vartheta_4(0|\tau);$$

$$(8.15) \quad \vartheta_2\left(0 \left| \frac{-1}{\tau + 1} \right. \right) = (-i(\tau + 1))^{1/2} \vartheta_3(0|\tau);$$

$$(8.16) \quad \vartheta_2(0|\tau) = 2e^{\pi i \tau / 4} + 2e^{9\pi i \tau / 4} + \dots;$$

$$(8.17) \quad \vartheta_4(0|\tau) = 1 - 2e^{\pi i \tau} + \dots$$

From (8.11), (8.12), and (8.13) we see that  $F(\tau)$  satisfies a transformation equation of the type (1.11) with  $r=1/2$  and

$$(8.21) \quad \epsilon = \epsilon(a, b, c, d) = \left( \frac{\pm c}{\pm d} \right) \exp \left\{ -\frac{\pi i}{4} (bd \pm 5d - 5 \pm 1) \right\}, \quad \pm d > 0,$$

for all transformations of the subgroup. By (8.16) we have

$$(8.22) \quad F(\tau) = (2e^{\pi i \tau / 4} + 2e^{9\pi i \tau / 4} + \dots)^{-1} = e^{7\pi i \tau / 4} (\frac{1}{2}e^{-2\pi i \tau} - \frac{1}{2} + \dots)$$

and by (8.14) and (8.17) we have

$$(8.23) \quad F(-1/\tau) = (-i\tau)^{-1/2} (1 + 2e^{\pi i \tau} + \dots).$$

On the other hand, (7.6) and (7.53) with  $A=D=1$ ,  $B=C=0$  yield

† See, for example, J. Tannery and J. Molk, *Fonctions Elliptiques*, vol. 2, 1896, p. 262.

$$\begin{aligned}
 F(\tau) &= \epsilon(1, -1, 2, -1) e^{-\pi i \alpha_1 \tau} e^{2\pi i \alpha_2 \tau} 2^{-1/2} f_2(e^{-\pi i} e^{2\pi i \tau}) \\
 (8.24) \quad &= 2^{-1/2} e^{-\pi i \alpha_1 \tau} e^{2\pi i \alpha_2 \tau} \sum_{m=-\mu_2}^{\infty} (-1)^m a_m^{(2)} e^{2\pi i m \tau},
 \end{aligned}$$

and (7.6) and (7.51) with  $A=D=0$ ,  $B=-1$ ,  $C=1$  give

$$(8.25) \quad F(-1/\tau) = (-i\tau)^{-1/2} e^{\pi i \alpha_1 \tau} \sum_{m=-\mu_1}^{\infty} a_m^{(1)} e^{\pi i m \tau}.$$

Comparing (8.22) with (8.24) and (8.23) with (8.25), we find

$$(8.26) \quad \alpha_1 = 0, \quad \alpha_2 = 7/8, \quad \mu_1 = 0, \quad \mu_2 = 1, \quad a_{-1}^{(2)} = 2^{-1/2} e^{-\pi i/8}, \quad a_0^{(1)} = 1.$$

Using these values we find that equations (7.41) and (7.43) now reduce to

$$(8.31) \quad a_n^{(1)} = 2^{-5/2} \pi e^{-\pi i/8} \sum_{\substack{k \equiv 1 \pmod{2} \\ k > 0}} \frac{1}{k} A_{k,1}^{(1)}(n) \frac{1}{n^{3/4}} I_{3/2}\left(\frac{\pi n^{1/2}}{k}\right), \quad n \geq 1,$$

where

$$\begin{aligned}
 (8.32) \quad A_{k,1}^{(1)}(n) &= \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \epsilon\left(h', \frac{h'k-1}{2h}, 2h'+2h, \frac{h'k-1}{h} + k\right)^{-1} \\
 &\quad \cdot \exp\left\{-\frac{\pi i}{8}\left(\frac{h'k-1}{kh} + 1\right) - 2\pi i n \frac{h}{k}\right\}, \quad k \equiv 1 \pmod{2},
 \end{aligned}$$

with  $h'$  defined by (7.31) and the  $\epsilon$  of (8.21). Similarly (7.48), (7.46), and (7.47) reduce to

$$\begin{aligned}
 (8.33) \quad a_n^{(2)} &= 2^{-7/4} \pi e^{-\pi i/8} \sum_{\substack{l \equiv 0 \pmod{2} \\ l > 0}} \frac{1}{l} A_{l',1}^{(2)}(n) \frac{1}{(n+7/8)^{3/4}} \\
 &\quad \cdot I_{3/2}\left(\frac{2\pi(n+7/8)^{1/2}}{l^{1/2}}\right), \quad n \geq 0,
 \end{aligned}$$

where  $l'$  is given by (7.49) and

$$\begin{aligned}
 (8.34) \quad A_{k,1}^{(2)}(n) &= \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \epsilon\left(h^{iv}, \frac{(k/2-h)h^{iv}-1}{2h}, 2h^{iv}+2h, \frac{(k/2-h)h^{iv}-1}{h} + \frac{k}{2} - h\right)^{-1} \\
 &\quad \cdot \exp\left\{-\frac{7\pi i}{4} \frac{h}{k} - \frac{\pi i}{8}\left(\frac{h^{iv}k-2}{kh} + 1\right) - 2\pi i n \frac{h}{k}\right\}, \quad \text{if } k \equiv 0 \pmod{4},
 \end{aligned}$$

$$\begin{aligned}
 & A_{k,1}^{(2)}(n) \\
 &= \sum_{\substack{0 \leq h < k \\ (h,k)=1}} \epsilon \left( h''', \frac{(k-2h)h'''-1}{4h}, 2h''' + 4h, \frac{(k-2h)h'''-1}{2h} - 2h + k \right)^{-1} \\
 & \cdot \exp \left\{ -\frac{7\pi i}{4} \frac{h}{k} - \frac{\pi i}{8} \left( \frac{h'''k-1}{2kh} + 1 \right) - 2\pi i n \frac{h}{k} \right\}, \text{ if } k \equiv 1 \pmod{2},
 \end{aligned}
 \tag{8.35}$$

with  $h'''$  and  $h^{iv}$  defined by (7.33) and (7.34) and the  $\epsilon$  of (8.21).

The Bessel functions which occur in (8.31) and (8.33) are of order half an odd integer and hence may be expressed in terms of elementary functions. Doing this we find that these equations may be written in the form

$$(8.36) \quad a_n^{(1)} = \frac{1}{2\pi} e^{-\pi i/8} \sum_{\substack{k \equiv 1 \pmod{2} \\ k > 0}} k^{1/2} A_{k,1}^{(1)}(n) \frac{d}{dn} \left( \frac{\sinh \frac{\pi}{k} n^{1/2}}{n^{1/2}} \right)$$

and

$$(8.37) \quad a_n^{(2)} = \frac{1}{2\pi} e^{-\pi i/8} \sum_{\substack{l \equiv 0 \pmod{2} \\ l > 0}} l^{1/2} A_{l',1}^{(2)}(n) \frac{d}{dn} \left( \frac{\sinh \frac{\pi}{l} (2(n+7/8))^{1/2}}{(n+7/8)^{1/2}} \right).$$

The expansion of  $F((A\tau+B)/(C\tau+D))$ , for  $\tau' = (A\tau+B)/(C\tau+D)$  any modular transformation, can now be obtained from (7.6). However we shall consider only three particular cases from which we shall get expansions for  $\vartheta_2(0|\tau)^{-1}$ ,  $\vartheta_3(0|\tau)^{-1}$ , and  $\vartheta_4(0|\tau)^{-1}$ . Two of these are to be found from (8.24) and (8.25). We insert the values (8.26) and have

$$(8.41) \quad F(\tau) = 2^{-1/2} e^{-7\pi i/8} e^{7\pi i\tau/4} \sum_{m=-1}^{\infty} (-1)^m a_m^{(2)} e^{2\pi i m \tau}$$

and

$$(8.42) \quad F(-1/\tau) = (-i\tau)^{-1/2} \sum_{m=0}^{\infty} a_m^{(1)} e^{\pi i m \tau}.$$

The third expansion is obtained from (7.6) and (7.52) by taking  $A=0$ ,  $B=-1$ ,  $C=D=1$ , and the values (8.26). We then find

$$(8.43) \quad F\left(\frac{-1}{\tau+1}\right) = (-i(t+1))^{-1/2} \sum_{m=0}^{\infty} (-1)^m a_m^{(1)} e^{\pi i m \tau}.$$

A comparison of (8.11) with (8.41), (8.15) with (8.43), and (8.14) with (8.42) then yields the desired expansions

$$(8.51) \quad \vartheta_2(0|\tau)^{-1} = 2^{-1/2} e^{-7\pi i/8} e^{7\pi i\tau/8} \sum_{m=-1}^{\infty} (-1)^m a_m^{(2)} e^{2\pi i m \tau},$$

$$(8.52) \quad \vartheta_3(0|\tau)^{-1} = \sum_{m=0}^{\infty} (-1)^m a_m^{(1)} e^{\pi i m \tau},$$

$$(8.53) \quad \vartheta_4(0|\tau)^{-1} = \sum_{m=0}^{\infty} a_m^{(1)} e^{\pi i m \tau}.$$

UNIVERSITY OF PENNSYLVANIA,  
PHILADELPHIA, PA.

# THE TYPE OF CERTAIN BOREL SETS IN SEVERAL BANACH SPACES†

BY

C. RAYMOND ADAMS AND JAMES A. CLARKSON

1. **Introduction.** Several writers have lately interested themselves in Borel sets in abstract spaces; especially in linear Borel sets in Banach spaces. In particular Mazur and Sternbach‡ have shown that a linear§  $F_{\sigma\delta}$  need not be an  $F_{\sigma}$ . To prove this they resorted to a somewhat elaborate construction of sets having the desired properties instead of selecting a familiar subset of a space already much studied; and whether their examples are  $G_{\delta\sigma}$ 's was not then determined. Shortly thereafter, however, Banach and Mazur|| gave two theorems concerning the Borel character of the convergence set (necessarily linear) of a sequence of linear operations, by aid of which they established that in every infinitely many-dimensional Banach space¶ there exists a linear  $F_{\sigma\delta}$  which is not a  $G_{\delta\sigma}$ ; and these theorems serve to fix completely the classification of some of the earlier examples.

Still more recently Oxtoby†† has examined, in the Lebesgue spaces‡‡  $L_p([0, 1])$ , ( $p \geq 1$ ), the set  $C$  of points corresponding to continuous functions and the set  $R$  corresponding to (properly) Riemann integrable functions. He showed that each of these sets is an  $F_{\sigma\delta}$  of first category; but he left unanswered, for example, the question of whether they are  $F_{\sigma}$ 's.

Following Oxtoby we consider here several familiar subsets of well known function spaces, and complete the determination of their Borel type. The sets

† Presented to the Society, September 6, 1938; received by the editors April 21, 1938.

‡ Mazur and Sternbach, *Über die Borelschen Typen von linearen Mengen*, *Studia Mathematica*, vol. 4 (1933), pp. 48-53.

§ Following Hausdorff we shall call a closed [open] set an  $F$  [ $G$ ]; the sum [product] of countably many closed [open] sets an  $F_{\sigma}$  [ $G_{\delta}$ ]; the product [sum] of countably many  $F_{\sigma}$ 's [ $G_{\delta}$ 's] an  $F_{\sigma\delta}$  [ $G_{\delta\sigma}$ ]; and so on.  $E$  is a set of order  $\alpha$  ( $\geq 0$ ) in the  $F$ -classification [ $G$ -classification] if it is an  $F$  [ $G$ ] with  $\alpha$  subscripts. Since every  $F$  [ $G$ ] is a  $G_{\delta}$  [ $F_{\sigma}$ ], any set of order  $\alpha$  is ambiguous of order  $\alpha+1$ ; that is, of order  $\alpha+1$  in both classifications. A set of order  $\alpha$  in both classifications, but of order  $\alpha-1$  in neither, may be called properly ambiguous of order  $\alpha$ .

By definition, a set  $E$  contained in a space  $S$  is of first category in  $S$  if it is the sum of countably many sets each non-dense in  $S$ ; otherwise it is of second category in  $S$ .

|| Banach and Mazur, *Eine Bemerkung über die Konvergenzmengen von Folgen linearer Operationen*, *Studia Mathematica*, vol. 4 (1933), pp. 90-94.

¶ In a finitely many-dimensional Banach space a linear set is always closed.

†† Oxtoby, *The category and Borel class of certain subsets of  $L_p$* , *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 245-248.

‡‡ The notation  $[a, b]$  will always designate the closed interval  $a \leq t \leq b$ .

$C$  and  $R \subset L_p$  will be shown to be unambiguous of order 2; the same will be proved for the set  $R^* \subset L_p$  corresponding to functions each of which is properly or improperly Riemann integrable<sup>†</sup> over  $[0, 1]$ , and for the set  $AC$  of absolutely continuous functions in the space  $C$  of continuous functions. The chief technical tool employed is Lemma 1, a characterization of  $G_\delta$  sets in any metric space (complete or not) which seems easy to apply in many particular instances. One of the above mentioned theorems of Banach and Mazur can also be used. Because of the lack of established technique for dealing with questions of the sort considered here we feel that quite as much interest (if not more) may attach to our methods as to our results. Therefore we offer no apology for developing our argument in such form that some of the results are proved more than once; the course of reasoning we shall follow seems the more clearly to illustrate the methods employed and the more accurately to indicate their range and ease of application.

It may be of interest that in  $L_p$  the set  $CBV$  corresponding to continuous functions of bounded variation is properly ambiguous of order 2.

We conclude this paper with a few remarks concerning other similar questions, including an illustrative application of our methods to a non-Banach space.

2. **Preliminary theorems.** Oxtoby has communicated to us the following theorem, of which he was in possession at the time his paper cited above was written; the simple proof given here, however, he obtained somewhat later.

**THEOREM 1 (Oxtoby).**  $R^*$  is an  $F_{\sigma\delta}$  of first category in each space  $L_p$ , ( $p \geq 1$ ).

**Proof.** If  $x, x'$  are arbitrary points of  $L_p$  and  $x(t), x'(t)$  any functions representative of these points, we have<sup>‡</sup>  $|x_N(t) - x'_N(t)| \leq |x(t) - x'(t)|$  for all  $t$  in  $[0, 1]$  and all  $N > 0$ ; hence  $T_N(x) = x_N$  is a continuous transformation of  $L_p$  into a part of itself. Setting  $R_N = T_N^{-1}(R)$ , we may write  $R^* = \bigcap_{N=1}^{\infty} R_N$ . Each  $R_N$ , being the antecedent of an  $F_{\sigma\delta}$  under  $T_N$ , is itself<sup>§</sup> an  $F_{\sigma\delta}$ ; whence  $R^*$  is an  $F_{\sigma\delta}$ . That  $R^*$  is of first category follows at once<sup>||</sup> since it is a proper linear Borel subset of  $L_p$ .

**THEOREM 2.**  $AC$  is an  $F_{\sigma\delta}$  of first category in the space  $C$ .

<sup>†</sup> We shall regard a function  $x(t) \in L_1$  unbounded on  $[0, 1]$  as *improperly* Riemann integrable if and only if for each  $N > 0$  the truncated function  $x_N(t)$  defined as  $-N, x(t)$ , or  $N$  according as  $x(t) < -N, -N \leq x(t) \leq N$ , or  $N < x(t)$  is R-integrable and  $\lim_{N \rightarrow \infty} \int_0^1 x_N(t) dt$  exists. With this definition,  $R^*$  is the set of points in  $L_p$  corresponding to functions which are Lebesgue integrable and continuous almost everywhere.

<sup>‡</sup> Here  $x_N(t)$  stands for the truncated function  $x(t)$  as defined in footnote <sup>†</sup> of this issue.

<sup>§</sup> See, for example, Kuratowski, *Topologie I*, Warsaw, 1933, p. 179.

<sup>||</sup> See Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 36, Theorem 1.

**Proof.** Let  $E_{mn}$ , ( $m, n=1, 2, 3, \dots$ ), represent the set of all continuous functions  $x(t)$  such that for any set of nonoverlapping subintervals  $t_r < t' < t'_r$  of  $[0, 1]$  with  $\sum_r (t'_r - t_r) < 1/m$  the condition  $\sum_r |x(t'_r) - x(t_r)| \leq 1/n$  is satisfied. For any fixed set of such subintervals, the condition imposed defines a closed set of points  $x \in C$ , and the product of any number of closed sets is closed; hence each  $E_{mn}$  is closed. But we clearly have  $AC = \prod_{n=1}^{\infty} \sum_{m=1}^{\infty} E_{mn}$ . That  $AC$  is of first category follows from Banach's theorem (loc. cit.).

**3. Concerning the  $F_\sigma$  property.** In this section we shall show that the sets mentioned in §1 are not  $F_\sigma$ 's in the spaces in question. We give first the following proof that the set  $C$  is not an  $F_\sigma$  in  $L_p$ .

It suffices to exhibit a set  $S \subset C$  such that under every decomposition  $S = \sum_{n=1}^{\infty} S_n$  at least one subset  $S_n$  has a limit point in  $L_p - C$ . Let  $CS$  designate the class of continuous singular functions,†  $DBVN$  the class of discontinuous functions of bounded variation having no external saltus.‡ According to recent results of Adams and Morse,§ if we introduce in the space  $BV$  of functions of bounded variation the metric||

$$(x, y) = \int_0^1 |x(t) - y(t)| dt + |T_0^1(x) - T_0^1(y)|,$$

$CS$  is a set of second category in  $CS + DBVN$ . Since  $DBVN$  is dense in the sum set,  $CS$  is not an  $F_\sigma$  therein; that is to say, under every decomposition  $CS = \sum_{n=1}^{\infty} H_n$  at least one set  $H_n$  has a limit point in  $DBVN$ . Now convergence in the metric of  $BV$  implies convergence in the metric of  $L_1$ , as well as uniform boundedness of the sequence of functions involved. Hence it follows that if  $S$  is the set of points in  $L_p$  corresponding to continuous singular functions and  $S = \sum_{n=1}^{\infty} S_n$  is any decomposition whatever, at least one subset  $S_n$  will have (in the metric of  $L_p$  for all  $p \geq 1$ ) a limit point in the set of points of  $L_p$  corresponding to  $DBVN$ . But a discontinuous function of bounded variation having no external saltus is not equivalent (in the metric of  $L_p$ ) to a continuous function; consequently the subset  $S_n$  in question has a limit point in  $L_p - C$ .

It may be observed that this reasoning shows that any set  $E$  such that  $CS \subset E \subset C$  in  $L_p$  is no  $F_\sigma$  in  $L_p$ . Moreover a precisely similar argument, in

† A singular function is a function of bounded variation whose derivative vanishes almost everywhere; see, for example, Saks, *Théorie de l'Intégrale*, Warsaw, 1933, pp. 11 ff.

‡ A function  $x(t)$  of bounded variation is said to have no external saltus if for every  $t_1$ , ( $0 \leq t_1 \leq 1$ ), we have  $\liminf_{t \rightarrow t_1} x(t) \leq x(t_1) \leq \limsup_{t \rightarrow t_1} x(t)$ .

§ Adams and Morse, *On the space (BV)*, these Transactions, vol. 42 (1937), pp. 194-205; see especially the concluding paragraph of §3.

|| Throughout this paper the distance between two points  $x, y$  in a metric space will be denoted by  $(x, y)$ . Here  $T_0^1(z)$  stands for the total variation of  $z(t)$  on the interval  $0 \leq t \leq 1$ .

which the space  $BV$  is metrized with the distance function†

$$(x, y) = \int_0^1 |x(t) - y(t)| dt + |L_0^1(x) - L_0^1(y)|$$

and  $CS$  is replaced by  $AC$ , can be employed to establish that any set  $E$  such that  $AC \subset E \subset C \subset L_p$  is no  $F_\sigma$  in  $L_p$ .

This type of proof, however, seems to be inapplicable to the other questions with which we are concerned. A method of attack with a much wider range of applicability is provided by the following lemma, which characterizes, by a property of the set itself, any  $G_\delta$  in a metric space.

**LEMMA† 1.** Let  $E$  be a subset of a metric space  $S$ . A necessary and sufficient condition that  $E$  be a  $G_\delta$  in  $S$  is the existence of a sequence of positive functions  $\Delta_n(x)$ , ( $n=1, 2, 3, \dots$ ), defined on  $E$  and having the property that no sequence  $\{x_n\} \subset E$  with  $(x_n, x_{n+1}) < \Delta_n(x_n)$  for all  $n$  converges to a point of  $S-E$ .

**Proof.** We consider first the necessity, and prove that if  $E$  is a  $G_\delta$ , there exists a sequence of functions  $\Delta_n(x)$  with the property asserted; in fact, each function  $\Delta_n(x)$  exhibited will in addition satisfy a Lipschitz condition of order 1 on  $E$ . Let  $E = \bigcap_{i=1}^\infty O_i$ , where  $O_i$  is open and  $O_{i+1} \subset O_i$  for each  $i$ ; let  $\rho(x, \bar{O}_i)$ , ( $x \in E$ ;  $i=1, 2, 3, \dots$ ), stand for the distance from  $x$  to the set  $\bar{O}_i = S - O_i$ ; and set  $\Delta_n(x) = 2^{-2n} \rho(x, \bar{O}_n)$ . If  $\{x_n\} \subset E$  satisfies the condition  $(x_n, x_{n+1}) < \Delta_n(x_n)$  for each  $n$ , we have for all  $s$

$$\Delta_{s+1}(x_{s+1}) \leq 2^{-2s-2} [(x_{s+1}, x_s) + \rho(x_s, \bar{O}_{s+1})] < 2^{-2s-1} \rho(x_s, \bar{O}_s) = \Delta_s(x_s)/2.$$

Hence for  $k$  fixed and  $n > k$  we infer

$$\rho(x_n, \bar{O}_k) \geq \rho(x_k, \bar{O}_k) - \sum_{s=k}^{n-1} (x_s, x_{s+1}) > \rho(x_k, \bar{O}_k) - 2\Delta_k(x_k) \geq \rho(x_k, \bar{O}_k)/2.$$

Therefore  $x_n$  cannot converge to a point of  $\bar{O}_k$  for any  $k$ ; hence it can converge to no point of  $\sum_{k=1}^\infty \bar{O}_k = S-E$ .

For the sufficiency, we note first that if there exists a sequence of functions  $\Delta_n(x)$  with the property specified, there certainly will exist one which has the additional properties  $\Delta_{n+1}(x) \leq \Delta_n(x) < 1/n$  for all  $x \in E$ , all  $n$ . Assuming then the existence of a sequence  $\Delta_n(x)$  with all these properties, we may set  $E_n = \sum_{x \in E} K(x, \Delta_n(x))$ , where  $K(x, \Delta_n(x))$  stands for the open sphere of  $S$  with center  $x \in E$  and radius  $\Delta_n(x)$ . Then each  $E_n$  is clearly open and  $E \subset \prod_{n=1}^\infty E_n$ . We shall show  $E \supset \prod_{n=1}^\infty E_n$ ; whence  $E = \prod_{n=1}^\infty E_n$  is a  $G_\delta$ . Let

† Here  $L_0^1(z)$  designates the (Peano) length of  $z(t)$  on the interval  $0 \leq t \leq 1$ . The set  $AC$  is of second category in  $AC+DBVN$ ; see Adams and Morse, loc. cit., the last paragraph of p. 204.

‡ This lemma, as well as its proof, is due entirely to Dr. Clarkson.—C.R.A.

$z \in \prod_{n=1}^{\infty} E_n$ ; then  $z \in E_1$  and there exists  $x_1 \in E$  such that  $(z, x_1) < \Delta_1(x_1)$ . Moreover, for every  $n$  there exists a point  $x'_n \in E$  with  $(z, x'_n) < \Delta_n(x'_n) < 1/n$ ; the sequence  $\{x'_n\}$  tends to  $z$ , and for  $n$  sufficiently large we have  $(x_1, x'_n) < \Delta_1(x_1)$ . Let  $x'_n$  (with  $n$  sufficiently large, and greater than or equal to 2) be taken as  $x_2$ ; then we have  $(z, x_2) = (z, x'_n) < \Delta_n(x'_n) = \Delta_n(x_2) \leq \Delta_2(x_2)$ . Next choose from  $\{x'_n\}$  an element such that  $(x_2, x'_n) < \Delta_2(x_2)$  with  $n \geq 3$  and call it  $x_3$ . Continuing indefinitely this process of selection we obtain  $\{x_n\}$ , a subsequence of  $\{x'_n\}$ , with  $(x_n, x_{n+1}) < \Delta_n(x_n)$  for every  $n$  and  $x_n \rightarrow z$ . The property of  $\Delta_n(x)$  stated in the lemma being assumed, we have  $z \in E$ .

**THEOREM 3.** *The sets  $C$ ,  $R$ , and  $R^*$  in each space  $L_p$ , ( $p \geq 1$ ), and the set  $AC$  in the space  $C$ , are no  $F_\sigma$ 's.*

**Proof.** We begin with a second proof that  $C$  is not an  $F_\sigma$  in  $L_p$ , using Lemma 1 to show that  $L_p - C$  is not a  $G_\delta$ . The basic idea of that lemma is that a sequence of points in a  $G_\delta$  which converges *rapidly enough* cannot tend to a limit outside the  $G_\delta$ , the rapidity of the convergence being prescribed by the condition  $(x_n, x_{n+1}) < \Delta_n(x_n)$ . The following argument shows that, given *any sequence whatsoever* of positive functions  $\Delta_n(x)$  defined on  $L_p - C$ , there always exists a sequence  $\{x_n\} \subset L_p - C$ , converging to a point in  $C$ , with  $\|x_n - x_{n+1}\| < \Delta_n(x_n)$  for every  $n$ .

Let  $p$  be fixed, let  $\{t_n\}$  be any sequence satisfying the conditions  $1/2 = t_1 < t_2 < \dots < t_n < \dots$ ,  $t_n \rightarrow t_0 \leq 1$ , and define  $x_1(t)$  as the characteristic function of the interval  $[0, t_1]$ . In general,  $x_n(t)$  having been defined, let  $x_{n+1}(t) = x_n(t)$  for  $t$  not in the interval  $[t_n, t_{n+1}]$ ; in that interval let  $x_{n+1}(t)$  be linear, with  $x_{n+1}(t_i) = 2^{1-i}$ , ( $i = n, n+1$ ). That each  $x_n \in L_p - C$  and that  $x_n \rightarrow x \in C$  is apparent. Moreover, it is clear that  $t_1, t_2, \dots, t_n$  having been fixed, we have  $\lim_{t_{n+1} - t_n} \|x_n - x_{n+1}\| = 0$ , so that by a proper selection of the sequence  $\{t_n\}$  the condition  $\|x_n - x_{n+1}\| < \Delta_n(x_n)$  may be satisfied for every  $n$ . This completes the proof for  $C \subset L_p$ . Actually  $x \in AC$ , so we have also proved that  $AC$  is no  $F_\sigma$  in  $L_p$ .

By a precisely similar argument we may dispose of  $R$  and  $R^*$  together, showing that for any sequence of positive functions  $\Delta_n(x)$  defined on  $L_p - R^*$  there exists a sequence  $\{x_n\} \subset L_p - R^*$ , with  $x_n \rightarrow x \in R$  and  $\|x_n - x_{n+1}\| < \Delta_n(x_n)$  for each  $n$ . Let  $y(t)$  be a function in the class†  $L_p$ , with  $|y(t)| < 1$  for all  $t$ , which in no interval  $[2^{-n-1}, 2^{-n}]$  is equivalent to a function in class  $R^*$ . We define  $x_1(t) = 0$  for  $0 \leq t \leq 1/2$ ,  $x_1(t) = y(t)$  otherwise; and in general,  $x_n(t)$  having been defined, set  $x_{n+1}(t) = 0$  for  $0 \leq t < 2^{-n-1}$ ,  $x_{n+1}(t) = k_{n+1}y(t)$  for  $2^{-n-1} \leq t < 2^{-n}$ , define  $x_{n+1}(t)$  as a continuous function in absolute value always less than 1 approximating  $x_n(t)$  in the norm of  $L_p$  for  $2^{-n} \leq t < 2^{-n+1}$ , and set

† We employ the term *class*  $L_p$  when an element is to be thought of as a single function.

$x_{n+1}(t) = x_n(t)$  for  $2^{-n+1} \leq t \leq 1$ , where  $\{k_n\}$  is a sequence of numbers with  $0 < k_n < 1$  for each  $n$ . This time it is clear that every  $x_n \in L_p - R^*$ ; and  $x_n \rightarrow x \in R$ , since  $x(t)$  is bounded and continuous almost everywhere. At each step of the process of definition the condition  $\|x_n - x_{n+1}\| < \Delta_n(x_n)$  will be satisfied if we merely choose the constant  $k_{n+1}$  sufficiently small and the continuous function in question to approximate  $x_n(t)$  closely enough in the norm of  $L_p$ .

That  $AC$  is no  $F_\sigma$  in space  $C$  may be established by a proof that so closely follows the line of the above argument that we leave its details to the reader, remarking only that the construction may conveniently be based on functions of the Cantor ternary type.

4. **Concerning the  $G_\delta$  property.** Our main object here is to establish the following theorem.

**THEOREM 4.** *The sets  $C$ ,  $R$ , and  $R^*$  in each space  $L_p$ , ( $p \geq 1$ ), and the set  $AC$  in the space  $C$ , are no  $G_\delta$ 's.*

For  $AC$  in space  $C$  a proof may be constructed by aid of this lemma:

**LEMMA 2.** *Let  $E$  be a subset of the space†  $AC$  which is dense in some sphere of  $AC$ ; then  $E$  contains a sequence which converges with arbitrary rapidity (in the sense specified by Lemma 1) in the metric of  $C$  to a point in  $C - AC$ .*

**Proof of the lemma.** On account of the homogeneity of  $AC$  as a Banach space, it suffices to consider the case in which  $E$  is dense in  $K$ , the unit sphere about the zero element as center. Let  $\Delta_n(x)$  be any sequence whatever of positive functions defined on  $E$ . Then  $x_1(t)$  may be taken as any element of  $E \cdot K$  satisfying the inequality  $1/2 < T_0^1(x_1) < 1$ . It is apparent that there exists a polygonal function  $p_1(t) \in K$ , its graph consisting of segments alternately horizontal and otherwise, such that the norm in  $C$ ,  $\|x_1 - p_1\|_C$ , is less than  $\Delta_1(x_1)/2$ ;  $T_0^1(p_1)$  exceeds  $1/2$ ; and the sum of the projections on the  $t$ -axis of the segments of the graph that are not horizontal, say  $\sum_{v=1}^{m_1} (t_{2v}^{(1)} - t_{2v-1}^{(1)})$ , is less than  $1/2$ . By virtue of the density of  $E$  in  $K$ , there exists a function  $x_2(t) \in E \cdot K$  with  $\|p_1 - x_2\|_C < \Delta_1(x_1)/2$  and

$$\sum_{v=1}^{m_1} |x_2(t_{2v}^{(1)}) - x_2(t_{2v-1}^{(1)})| = 1/2 + \delta_1, \quad \delta_1 > 0.$$

Next we choose a polygonal function  $p_2(t) \in K$ , its graph consisting of segments alternately horizontal and sloping, such that  $\|x_2 - p_2\|_C$  is less than  $\min [\Delta_2(x_2)/2, \delta_1/(8m_1)]$ ;  $T_0^1(p_2)$  exceeds  $1/2$ ; and the sum of the projections on the  $t$ -axis of the sloping segments of the graph, say  $\sum_{v=1}^{m_2} (t_{2v}^{(2)} - t_{2v-1}^{(2)})$ , is

† The space  $AC$ , here understood to be normed with  $\|x\| = |x(0)| + T_0^1(x)$ , is a Banach space.

less than  $1/2^2$ . Then we select a function  $x_3(t) \in E \cdot K$  with  $\|p_2 - x_3\|_C < \min [\Delta_2(x_2)/2, \delta_1/(8m_1)]$  and

$$\sum_{\nu=1}^{m_2} |x_3(t_{2\nu}^{(2)}) - x_3(t_{2\nu-1}^{(2)})| = 1/2 + \delta_2, \quad \delta_2 > 0.$$

In the third stage of this procedure  $p_3(t) \in K$  would be chosen to satisfy the inequalities

$$\|x_3 - p_3\|_C < \min [\Delta_3(x_3)/2, \delta_1/(16m_1), \delta_2/(8m_2)], \quad T_0^1(p_3) > 1/2,$$

$$\sum_{\nu=1}^{m_3} (t_{2\nu}^{(3)} - t_{2\nu-1}^{(3)}) < 1/2^3;$$

and  $x_4(t) \in E \cdot K$  to fulfill the conditions

$$\|p_3 - x_4\|_C < \min [\Delta_3(x_3)/2, \delta_1/(16m_1), \delta_2/(8m_2)],$$

$$\sum_{\nu=1}^{m_3} |x_4(t_{2\nu}^{(3)}) - x_4(t_{2\nu-1}^{(3)})| = 1/2 + \delta_3, \quad \delta_3 > 0.$$

And so we continue for subsequent stages, indefinitely. It is quite clear that  $\{x_n\}$  is a Cauchy sequence in the metric of space  $C$ , and that if  $x$  is its limit, we have  $\sum_{\nu=1}^{m_k} |x(t_{2\nu}^{(k)}) - x(t_{2\nu-1}^{(k)})| > 1/2$  for each integer  $k \geq 1$ ; since  $\sum_{\nu=1}^{m_k} (t_{2\nu}^{(k)} - t_{2\nu-1}^{(k)}) \rightarrow 0$  with  $1/k$ , we have  $x \in C - AC$ , and the proof of the lemma is complete.

Returning now to the proof of Theorem 4 for  $AC$  in space  $C$ , we observe that since the space  $AC$  is complete and consequently of second category in itself, under any decomposition whatsoever  $AC = \sum_{n=1}^{\infty} E_n$  at least one set  $E_n$  must be dense (in the metric of  $AC$ ) in some sphere of  $AC$ . By Lemma 2 we conclude that this set  $E_n$  contains a sequence of the sort specified in the statement of that lemma. Hence, by Lemma 1,  $E_n$  is no  $G_\delta$  in  $C$ ; and  $AC$  is no  $G_\delta$ .

For  $C$ ,  $R$ , and  $R^*$  in space  $L_p$  one may employ an entirely similar proof using the following lemma.

**LEMMA 3.** *Let  $E$  be a subset of the space  $C$  which is dense in some sphere of  $C$ ; then  $E$  contains a sequence which converges with arbitrary rapidity (in the sense specified by Lemma 1) in the metric of  $L_p$ , ( $p \geq 1$ ), to a point in  $L_p - R^*$ .*

**Proof of the lemma.** On account of the homogeneity of  $C$  as a Banach space, it is sufficient to consider the case in which  $E$  is dense in  $K$ , the unit sphere about the zero element as center. Let  $\Delta_n(x)$  be any sequence whatever of positive functions defined on  $E$ . A sequence of closed subintervals  $\delta_j$ , ( $j=1, 2, 3, \dots$ ), of  $I=[0, 1]$  is to be chosen as follows:  $\delta_1$  is concentric with  $I$ ;  $\delta_2$  and  $\delta_3$  are concentric with the two intervals to the left and right, respectively, of  $\delta_1$ ; the next four  $\delta_j$  are concentric, respectively, with the four

subintervals of  $I$  which constitute the point set  $I - (\delta_1 + \delta_2 + \delta_3)$ ; and so on, so that the point set  $\sum_{j=1}^{\infty} \delta_j$  is dense in  $I$ . For each  $j$  we allow  $\delta'_j$  to stand for the closed middle third of  $\delta_j$ .

At the outset the condition  $m(\sum_{j=1}^{\infty} \delta_j) < 1$  will be imposed.

As a first step in the construction of the required sequence of functions  $x_n(t)$ , let  $\delta_1$  be chosen arbitrarily and  $x_1(t) \in E \cdot K$  so that

$$x_1(t) > 3/4 \text{ for } t \in \delta'_1, \quad x_1(t) < 1/4 \text{ for } t \in I - \delta_1.$$

The density of  $E$  in  $K$  clearly insures the existence of such a function  $x_1(t)$ . Secondly, if  $\delta_2$  and  $\delta_3$  are small enough, there will exist a function  $x_2(t) \in E \cdot K$  with

$$x_2(t) > 3/4 \text{ for } t \in \delta'_1 + \delta'_2 + \delta'_3, \quad x_2(t) < 1/4 \text{ for } t \in I - (\delta_1 + \delta_2 + \delta_3),$$

$$(x_1, x_2) < \Delta_1(x_1),$$

where  $(x_1, x_2) = \|x_1 - x_2\|$  in  $L_p$ ; again this is insured by the density of  $E$  in  $K$ , and we assume  $x_2(t)$  to be so chosen. Let this procedure be continued indefinitely; at each stage the existence of sufficiently small intervals  $\delta_i$ , and of a suitable function  $x_n(t)$  to satisfy the desired conditions (whose nature must now be apparent to the reader), should be clear from the steps already described in detail. The imposition of a further condition, such as, for example,  $(x_n, x_{n+1}) < 2^{-n}$ , will insure that  $x_n(t)$  converge in  $L_p$  to some function  $x(t)$ .

For  $t \in H = I - \sum_{j=1}^{\infty} \delta_j$ , each function  $x_n(t)$  is less than  $1/4$ , and  $m(H)$  is greater than 0; hence  $x(t) \leq 1/4$  almost everywhere in  $H$ , and by aid of metric density we infer  $\text{ess lim inf}_{t \rightarrow t_1} x(t) \leq 1/4$  for almost all  $t_1 \in H$ . On the other hand, for  $t$  in any interval  $\delta'_j$ , all  $x_n(t)$  with  $n$  sufficiently large are greater than  $3/4$ , whence  $x(t) \geq 3/4$  for almost all  $t \in \sum_{j=1}^{\infty} \delta'_j$ ; since each point of  $H$  is a limit point of intervals  $\delta'_j$ , this implies  $\text{ess lim sup}_{t \rightarrow t_1} x(t) \geq 3/4$  everywhere in  $H$ . Thus there is a set of measure greater than zero at each point of which the essential saltus of  $x(t)$  is greater than or equal to  $1/2$ , and  $x \in L_p - R^*$ .

We may now show that any set  $E$  such that  $C \subset E \subset R^* \subset L_p$  is no  $G_\delta$  in  $L_p$ . Since space  $C$  is of second category in itself, under any decomposition whatever  $E = \sum_{n=1}^{\infty} E_n$  at least one of the sets  $C \cdot E_n$  contains a set of continuous functions which is dense (in the metric of  $C$ ) in some sphere of  $C$ . The conclusion follows from Lemmas 3 and 1.

Oxtoby (loc. cit.) has proved that the set  $S_u[S_i] \subset L_p$  of points corresponding to functions each of which is upper [lower] semicontinuous on  $[0, 1]$  is an  $F_\sigma$ , and has observed that  $R = S_u \cdot S_i$ . Since  $S_u$  and  $S_i$  are clearly of the same Borel type, the determination of their type is made precise by the following corollary.

COROLLARY. The sets  $S_n$  and  $S_1$  in each space  $L_p$ , ( $p \geq 1$ ), are no  $G_{\delta\sigma}$ 's.

For  $AC$  in space  $C$ , and  $C$  in space  $L_p$ , alternative proofs<sup>†</sup> of Theorem 4 can easily be made by means of Theorem 3 and the following result already spoken of in §1.

LEMMA 4 (Banach and Mazur). Let  $U_n(x)$ , ( $n=1, 2, 3, \dots$ ), be a sequence of operations each linear<sup>‡</sup> on a Banach space  $S$  to a like space  $S'$ , and let  $E$  be the subset of  $S$  at each point  $x$  of which  $U_n(x)$  converges in  $S'$ . If  $E$  is a  $G_{\delta\sigma}$ , it is an  $F_\sigma$ .

A sequence of linear operations on space  $C$  to  $L_1$  having  $AC$  for convergence set is provided by the following lemma (we choose  $U_n(x) = p_n'(t)$ ).

LEMMA 5. Let  $x \in C$  and let  $p_n(t)$  be the polygonal function inscribed in  $x(t)$  with "corners" at  $t=m/n$ , ( $n=1, 2, 3, \dots$ ;  $m=0, 1, \dots, n$ ). A necessary and sufficient condition that  $x \in AC$  is that the sequence  $p_n'(t)$  converge in the space  $L_1$ .

**Proof of the lemma.** For the necessity we have  $x \in AC$  and  $p_n(t)$  converging in length to  $x(t)$ ; by a theorem of Adams and Lewy<sup>§</sup> it follows that  $T_0^1(p_n - x) \rightarrow 0$ . But since  $p_n \in AC$  for each  $n$ , we have

$$T_0^1(p_n - x) = \int_0^1 |p_n'(t) - x'(t)| dt.$$

For the sufficiency we may assume without real restriction that  $x(0)=0$ . Then, if  $p_n'(t) \rightarrow y(t) \in L_1$ , we may set  $z(t) = \int_0^t y(s) ds$  and obtain for all  $t$  in  $[0, 1]$

$$|p_n(t) - z(t)| = \left| \int_0^t [p_n'(s) - y(s)] ds \right| \leq \int_0^t |p_n'(s) - y(s)| ds \rightarrow 0.$$

Since also  $|p_n(t) - x(t)| \rightarrow 0$  for all  $t$ , we infer  $x(t) \equiv z(t)$ .

For  $C \subset L_p$  a similar proof can be constructed. Let us consider first the case of  $p=1$ , and for each  $x \in L_1$ , each integer  $n \geq 1$ , define

$$U_n(x) = y_n(\xi) = \begin{cases} n \int_{\xi}^{\xi+1/n} x(t) dt & \text{for } 0 \leq \xi \leq 1 - 1/n, \\ y_n(1 - 1/n) & \text{for } 1 - 1/n \leq \xi \leq 1. \end{cases}$$

<sup>†</sup> The applicability of these alternative proofs, however, is limited to Banach spaces. Moreover one such proof cannot yield the desired result for an *entire range* of sets in a space, as does, for example, the above proof based on Lemma 3.

<sup>‡</sup> Linear in the sense of Banach; that is, additive and continuous.

<sup>§</sup> Adams and Lewy, *On convergence in length*, Duke Mathematical Journal, vol. 1 (1935), pp. 19-26, Theorem 4.

Then every  $U_n(x)$  is a linear operation on  $L_1$  to space  $C$ . If  $x \in C \subset L_1$ , and  $\xi$  is any point in the interval  $0 \leq \xi < 1$ , there exists  $\xi'$  with  $0 \leq \xi' - \xi \leq 1/n$  and  $y_n(\xi) = x(\xi')$ , so that the uniform continuity of a representative function  $x(t)$  implies  $y_n(t) \rightarrow x(t)$  uniformly on  $[0, 1]$ . On the other hand, if the sequence  $y_n(\xi)$  converges in space  $C$ , let  $y(\xi)$  be its limit; that  $y(\xi) = x(\xi)$  for almost all  $\xi$  is well known, and we have  $y \in C$ . Hence  $C \subset L_1$  is the convergence set of the sequence of linear operations  $U_n(x)$ . For the case of  $p > 1$  the same reasoning is valid,† the linearity of each operation being a consequence of the inequality  $\|U_n(x)\| \leq n^{1/p} \|x\|$ .

Success has not attended our efforts to find a sequence of linear operations on  $L_p$  having for its convergence set either  $R$  or  $R^*$ .

We conclude this section by exhibiting a properly ambiguous set of order 2 in  $L_p$ ; namely,  $CBV$ . That is, we shall establish the following theorem.

**THEOREM 5.** *The set  $CBV$  in each space  $L_p$ , ( $p \geq 1$ ), is simultaneously an  $F_{\sigma\delta}$  and a  $G_{\delta\sigma}$ , without being either an  $F_\sigma$  or a  $G_\delta$ .*

To prove this theorem we note first that  $BV \subset L_p$  may easily be proved‡ an  $F_\sigma$  by setting  $BV = \sum_{n=1}^{\infty} E_n$  where  $E_n$  is the set of points each of which has a representative function  $x(t)$  with  $T_0^1(x) \leq n$ . That each  $E_n$  is closed in  $L_p$  is an immediate consequence of a well known theorem of Helly.§

Now it is easily seen that in  $L_p$ ,  $CBV = C \cdot BV$ . Hence, in the light of Oxtoby's result (loc. cit.) that  $C$  is an  $F_{\sigma\delta}$ ,  $CBV$  is an  $F_{\sigma\delta}$ ; that it is no  $F_\sigma$  has been shown above in §3, and that it is no  $G_\delta$  follows from the theorem of Mazur and Sternbach just cited.‡ It therefore remains only for us to prove that  $CBV$  is a  $G_{\delta\sigma}$ . This will be done by aid of the following lemma.

**LEMMA 6.** *Let  $x(t)$ ,  $x_0(t)$  be elements of the class  $CBV$ , and let  $\|x - x_0\|_C$  and  $\|x - x_0\|_{L_p}$  represent the norms in the spaces  $C$  and  $L_p$ , ( $p \geq 1$ ), respectively; then the relation  $\|x - x_0\|_{L_p} \rightarrow 0$  implies*

$$\liminf (T_0^1(x) - \|x - x_0\|_C) \geq T_0^1(x_0).$$

**Proof of the lemma.** It suffices to establish the conclusion for the case of a

† Another proof for  $C \subset L_p$  can be based on the following simple lemma, which provides a sequence of linear operations ( $U_n(x) = s_n(t)$ ) on  $L_p$  to space  $M$  (or  $L_\infty$ ) having  $C$  for convergence set. Let  $x$  be an arbitrary element of  $L_p$ , and let  $s_n(t)$  be the step-function defined on each subinterval  $m/n \leq t < (m+1)/n$ , ( $n = 1, 2, 3, \dots$ ;  $m = 0, 1, \dots, n-1$ ), as the integral mean of  $x(t)$  on that subinterval,  $s_n(1) = s_n(1-0)$ , so that  $s_n(t)$  is an element of the class  $M$  (or  $L_\infty$ ) of essentially bounded measurable functions. Then a necessary and sufficient condition that  $x \in C$  is that the sequence  $s_n$  converge in the space  $M$  (normed, as usual, with  $\|s\| = \text{ess sup}_{0 \leq t \leq 1} |s(t)|$ ).

‡ Mazur and Sternbach (loc. cit.) have proved that a linear  $G_\delta$  in a Banach space is closed; hence  $BV$  is no  $G_\delta$ , and its Borel type is completely determined.

§ Helly, *Über lineare Funktionaloperationen*, Sitzungsberichte der Wiener Akademie, class IIa, vol. 121 (1912), p. 283.

sequence  $x_n(t)$ , ( $n=1, 2, 3, \dots$ ), with  $\|x_n - x_0\|_{L_p} \rightarrow 0$ . This result will follow at once if we show that  $x_n(t) \rightarrow x_0(t)$  on a set  $D$  of points  $t$  dense in  $[0, 1]$  implies the conclusion.

Let  $\epsilon$  be an arbitrary positive number and let  $S: 0=t_0 < t_1 < t_2 < \dots < t_k=1$  be a set of points with  $t_i \in D$  for  $i=1, 2, \dots, k-1$  and such that

$$T_{t_{i-1}}^{t_i}(x_0) < \epsilon \text{ for } i=1, 2, \dots, k; \quad \sum_{i=1}^k |x_0(t_i) - x_0(t_{i-1})| > T_0^1(x_0) - \epsilon.$$

Next let  $N$  be such that for all  $n > N$  we have

$$(1) \quad |x_n(t_i) - x_0(t_i)| < \epsilon/(2k), \quad i=1, 2, \dots, k-1;$$

whence

$$\sum_{i=2}^{k-1} |x_n(t_i) - x_n(t_{i-1})| > \sum_{i=2}^{k-1} |x_0(t_i) - x_0(t_{i-1})| - \epsilon > T_0^1(x_0) - 4\epsilon.$$

We now fix  $n$  as any integer greater than  $N$ . Let  $t'$  be a point where

$$\max_{0 \leq t \leq 1} |x_n(t) - x_0(t)|$$

is assumed. Then  $t'$  is contained in one of the subintervals determined by  $S$ , say in  $[t_{i-1}, t_i]$ ; and for at least one of the end points of this subinterval, say  $t_i$ , inequality (1) holds. Therefore we have

$$\begin{aligned} |x_n(t') - x_n(t_i)| &\geq |x_n(t') - x_0(t')| - |x_0(t') - x_0(t_i) + x_0(t_i) - x_n(t_i)| \\ &\geq \|x_n - x_0\|_C - 2\epsilon; \end{aligned}$$

whence

$$\begin{aligned} T_0^1(x_n) &\geq \left[ \sum_{i=2}^{l-1} + \sum_{i=l+1}^{k-1} \right] |x_n(t_i) - x_n(t_{i-1})| + |x_n(t') - x_n(t_i)| \\ &\geq \left[ \sum_{i=2}^{l-1} + \sum_{i=l+1}^{k-1} \right] |x_0(t_i) - x_0(t_{i-1})| - \epsilon + \|x_n - x_0\|_C - 2\epsilon \\ &\geq \sum_{i=2}^{k-1} |x_0(t_i) - x_0(t_{i-1})| + \|x_n - x_0\|_C - 4\epsilon \\ &\geq T_0^1(x_0) + \|x_n - x_0\|_C - 7\epsilon, \end{aligned}$$

and the proof of the lemma is complete.

Returning now to the proof of Theorem 5, we set  $CBV = \sum_{k=1}^{\infty} E_k$  where  $E_k$  is the set of points in  $CBV \subset L_p$  each of which has a representative function  $x(t)$  with  $T_0^1(x) \leq k$ . It suffices to show that each  $E_k$  is a  $G_\delta$  in  $L_p$ . Having fixed  $k$ , we define the function  $\Delta_n(x)$  specified in Lemma 1 as follows: for

each  $x \in E_k$ , each integer  $n \geq 1$ , let  $\Delta_n(x) = r > 0$  where  $r$  is such that for  $y \in E_k \cdot K(x, r)$  we have

$$\|x - y\|_C \leq T_0^1(y) - T_0^1(x) + 1/2^n,$$

whence

$$(2) \quad T_0^1(y) - T_0^1(x) \geq -1/2^n,$$

this  $r$  existing in consequence of Lemma 6. If  $x_n$ , ( $n = 1, 2, 3, \dots$ ), is any sequence contained in  $E_k$ , with  $\|x_{n+1} - x_n\|_{L_p} < \Delta_n(x_n)$ , there is a sequence of representative functions  $x_n(t)$  for which we have

$$\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|_C \leq \sum_{n=1}^{\infty} [T_0^1(x_{n+1}) - T_0^1(x_n)] + \sum_{n=1}^{\infty} 1/2^n.$$

The first series on the right is easily proved convergent, since its partial sum  $T_0^1(x_m) - T_0^1(x_1)$  is "almost non-decreasing" by virtue of inequality (2) and is bounded from above by  $2k$ . Hence  $x_n(t)$  converges uniformly, say to  $x(t) \in C$ ;  $T_0^1(x) \leq k$  by Lemma 6; the corresponding element  $x \in L_p$  to which  $x_n$  converges in  $L_p$  is in  $E_k$ ; and  $E_k$  is a  $G_\delta$  in  $L_p$  by Lemma 1.

**5. Other questions.** The above sections afford methods and illustrate techniques by means of which certain questions concerning the Borel character of many sets in various metric spaces can be settled. Without prolonging our discussion unduly we may list here, with slight indications of proof, the answers to a few questions which the reader might very naturally raise in view of the sets and spaces already considered.

- I. *AC in the spaces  $M$  and  $L_p$  is an  $F_{\sigma\delta}$ , no  $G_{\delta\sigma}$ .*
- II. *CBV in the spaces  $C$  and  $M$  is an  $F_\sigma$ , no  $G_\delta$ .*
- III. *BV in the space  $M$  is an  $F_\sigma$ , no  $G_\delta$ .*
- IV.  *$C$  in the space  $M$  is an  $F$ , no  $G$ .*
- V.  *$R$  in the space  $M$  is an  $F$ , no  $G$ .*

In each case the set is of first category in the space in question, according to the theorem of Banach cited earlier. In cases I-III the set is no  $G_\delta$  by virtue of the theorem of Mazur and Sternbach cited in §4. For determining the  $F$ -classification of  $AC$  in  $L_p$  the following observations may be helpful.

(i) Let  $x_n(t)$ , ( $n = 1, 2, 3, \dots$ ), be continuous on  $[0, 1]$  and  $x_n(t) \rightarrow x(t)$  in the norm of  $L_p$ . If for arbitrary  $\epsilon > 0$  there exists  $\delta > 0$  such that for every  $n$  we have

$$|x_n(t') - x_n(t'')| < \epsilon \quad \text{for} \quad |t' - t''| < \delta,$$

then this condition is satisfied also by  $x(t)$ .

(ii) For fixed  $m, n$ , ( $m, n = 1, 2, 3, \dots$ ), the functions  $x(t)$ , continuous on

$[0, 1]$  and such that for any set of nonoverlapping intervals  $t_r < t < t'_r$  with  $\sum_r (t'_r - t_r) < 1/m$  the condition  $\sum_r |x(t'_r) - x(t_r)| \leq 1/n$  is satisfied, constitute a closed set in the space  $C$  normed with the norm of  $L_p$ .

In connection with the proof that  $R$  is closed in space  $M$  it may be advantageous to note this fact:

(iii) If  $x(t)$  is an element of class  $M$  with the property that  $\text{ess } \lim_{t \rightarrow t_1} x(t)$  exists for almost all  $t_1 \in [0, 1]$ , there exists a function  $y(t)$  in class  $R$  which equals  $x(t)$  almost everywhere.

A considerable number of the listed results can be derived at once from others by means of the following simple observation concerning relativization:

(iv) If  $S$  is a metric space,  $A \subset B \subset S$ , and  $A$  a Borel set of a certain type relative to  $S$ , then  $A$  is a Borel set of the same type relative to the space  $B$  metrized with any metric in which convergence implies convergence in the metric of  $S$ .

In conclusion, we should like to emphasize that the range of applicability of the methods mainly employed above is by no means restricted to Banach spaces. By way of illustration, let us consider the set  $AC$  in the space  $BV$  metrized with the distance function mentioned in the second paragraph of §3, which is not a Banach space. Adams and Morse (loc. cit., p. 201) have observed that  $AC$  is no  $G_\delta$  relative to  $CBV \subset BV$ ; that it is no  $G_\delta$  in  $BV$  follows at once; that it is no  $G_{\delta\sigma}$  in  $BV$  may be shown easily as follows, although Lemma 4 (Banach and Mazur) is obviously not applicable. From Lemma 6 one may infer that the limit function  $x(t)$  determined in the proof of Lemma 2 satisfies the condition  $T_\delta^1(x) \leq \liminf_{n \rightarrow \infty} T_\delta^1(x_n)$ ; in the course of that proof one may clearly impose an additional condition on the rapidity of convergence of  $\|x_n - x_{n+1}\|_C$  to zero to insure  $T_\delta^1(x) > T_\delta^1(x_k) - 1/2^k$  for every  $k$ ; then  $x_n$  tends to  $x \in CBV - AC$  in the metric of  $BV$ . The reasoning contained in the first paragraph following the proof of Lemma 2 may now be duplicated to show that  $AC$  is no  $G_{\delta\sigma}$  in  $BV$ . That  $AC$  is an  $F_{\sigma\delta}$  in  $BV$  is an immediate consequence of I and observation (iv). Since  $CBV$  is already known to be a  $G_\delta$  in  $BV$ , it follows at once that  $AC$  is an  $F_{\sigma\delta}$  but no  $G_{\delta\sigma}$  relative to  $CBV \subset BV$ .

BROWN UNIVERSITY,  
PROVIDENCE, R. I.  
THE UNIVERSITY OF PENNSYLVANIA,  
PHILADELPHIA, PA.

## RESIDUATED LATTICES\*

BY

MORGAN WARD AND R. P. DILWORTH

### I. INTRODUCTION

1. We propose to develop here a systematic theory of lattices<sup>†</sup> over which an auxiliary operation of multiplication or residuation is defined. We begin by showing that the two operations correspond to one another; under quite general conditions in every lattice over which a multiplication is defined a residuation may be defined and conversely. The residuation and multiplication we introduce have the properties of the like-named operations in the particular instance of polynomial ideal theory.

We next give various necessary conditions and sufficient conditions that such operations may exist in an arbitrary lattice, and apply our results to projective geometries and Boolean algebras.

In the third division of the paper we extend E. Noether's decomposition theorems of the ideal theory of commutative rings to general lattice theory. The introduction of a multiplication is obviously necessary for such a generalization. The surprising result emerges that the decomposition theorems are largely independent of the modular axiom, as we show by specific examples. We take this occasion to correct an error made in the preliminary account of our researches (Ward and Dilworth [1]). Since we wrote this, we have obtained many new results which we give here for the first time.

We plan to describe the main part of our investigations of distributive residuated lattices elsewhere (Ward and Dilworth [1], §§5, 6). Here we settle some questions raised by one of us (Ward [1]) as to the significance of certain auxiliary conditions which a residuation may satisfy by showing in all cases that they imply that the lattice is distributive.

2. It was not until this paper was virtually completed that we learned of the investigation of Krull upon this subject (Krull [1]). There is, however, very little duplication between our results and Krull's. Krull was chiefly concerned with the problem of finding out in what manner the Noether decomposition theorems could be extended to a residuated lattice in which the chain condition was weakened and no connection was assumed between irreducibles and primary elements.

\* Presented to the Society, March 27, 1937, and April 9, 1938; received by the editors April 21, 1938.

† For a connected account of lattice theory and the literature up to 1937, see Köthe [1].

3. We shall use the following terminology and notation.  $\mathfrak{S}$  is a fixed lattice with elements  $a, \dots, y$  with or without subscripts. Sublattices of  $\mathfrak{S}$  are denoted by German capitals  $\mathfrak{A}, \mathfrak{B}$ . The letters  $\mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$  are reserved to denote subsets of  $\mathfrak{S}$  which are not necessarily sublattices. We write  $x \in \mathfrak{X}$  for "the set  $\mathfrak{X}$  contains the element  $x$ ." The expressions  $x \supset y$  or  $y \subset x$ ,  $x \nmid y$  denote, as usual,  $x$  divides  $y$ ,  $x$  does not divide  $y$ . We write  $x = y$  if  $x \supset y$  and  $y \supset x$  (Ore [1], p. 42) and  $x > y$  or  $y < x$  for  $x$  covers  $y$  (Birkhoff [1]). We use  $(x, y)$  and  $[x, y]$  for union and cross-cut. If the unit and null elements exist, we denote them by  $i$  and  $z$ , respectively. Elements covered by  $i$  are called divisor-free. If  $(a, b) = i$ ,  $a$  and  $b$  are said to be co-prime. If every pair of distinct elements of a set  $\mathfrak{X}$  are co-prime, the set is said to be co-prime. An element  $n$  of  $\mathfrak{S}$  is called a node if either  $x \supset n$  or  $n \supset x$  for every  $x$  of  $\mathfrak{S}$ . A sublattice  $\mathfrak{A}$  is said to be dense over  $\mathfrak{S}$  if  $a_1, a_2 \in \mathfrak{A}$  and  $a_1 \supset x \supset a_2$  imply  $\mathfrak{A}$  contains  $x$ . If every set of elements finite or infinite of  $\mathfrak{S}$  has a cross-cut (union),  $\mathfrak{S}$  is said to be completely closed relative to cross-cut (union). Two properties  $P$  and  $Q$  which  $\mathfrak{S}$  may possess are said to be completely independent if there exist instances of lattices in which both  $P$  and  $Q$  hold, neither holds,  $P$  holds but not  $Q$ ,  $Q$  holds but not  $P$ .

We shall find it convenient to use the following conditions for a distributive lattice either of which is equivalent to the usual formulation:

- (i)  $b \supset [a, c]$  implies  $b = [(b, a), (b, c)]$ .
- (ii)  $(a, c) \supset b$  implies  $b = [b, a], [b, c]$ .

## II. RESIDUATIONS AND MULTIPLICATIONS

4. Assume that  $\mathfrak{S}$  contains  $i$ . A well-defined one-valued binary operation  $x:y$  is called a residuation over  $\mathfrak{S}$  if the following conditions are satisfied:

- R 1. If  $\mathfrak{S}$  contains  $a, b$ , then  $\mathfrak{S}$  contains  $a:b$ .
- R 2.  $a:b = i$  if and only if  $a \supset b$ .
- R 3.  $a \supset b$  implies that  $a:c \supset b:c$  and  $c:b \supset c:a$ .
- R 4.  $(a:b):c = (a:c):b$ .
- R 5.  $[a, b]:c = [a:c, b:c]$ .
- R 6.  $c:(a, b) = [c:a, c:b]$ .

We postpone the consideration of the dual residuation for our second paper.

A well defined binary operation  $x \cdot y$  (or  $xy$ ) is called a multiplication over  $\mathfrak{S}$  if the following conditions are satisfied:

- M 1. If  $\mathfrak{S}$  contains  $a, b$ , then  $\mathfrak{S}$  contains  $a \cdot b$ .
- M 2. If  $a = b$ , then  $a \cdot c = b \cdot c$ .
- M 3.  $a \cdot b = b \cdot a$ .

$$M\ 4. (a \cdot b) \cdot c = a \cdot (b \cdot c).$$

$$M\ 5. \text{ If } \mathfrak{S} \text{ contains } i, \text{ then } a \cdot i = a.$$

$$M\ 6. a \cdot (b, c) = (a \cdot b, a \cdot c).$$

It may be shown (Ward [1]) that a residuation exists satisfying R 1-R 6 if a multiplication over  $\mathfrak{S}$  exists satisfying M 1-M 6 and the following condition:

M 7.  $\mathfrak{S}$  is completely closed with respect to union, and the product of the unions of any two sets of elements of  $\mathfrak{S}$  is the union of the products of all pairs of elements in the sets.<sup>1</sup>

This residual  $a:b$ , satisfying R 1-R 6, is defined as follows:

DEFINITION 4.1. (i)  $a \supset (a:b)b$ ; (ii) if  $a \supset xb$ , then  $a:b \supset x$ .

If we take for  $x \cdot y$  the cross-cut  $[x, y]$ , then conditions M 1-M 6 are all satisfied provided that  $\mathfrak{S}$  is distributive. If M 7 holds, the lattice is said to be completely distributive with respect to union. Hence (Ward [2]) every completely distributive lattice may be residuated in at least one way.

Another condition\* insuring the existence of a residual is the following:

M 8. For any two elements  $a, b$  of  $\mathfrak{S}$ , the ascending chain condition holds in the set  $\mathfrak{X}$  of all  $x$  such that  $a \supset xb$ .

M 8 insures the existence of a union of the set  $\mathfrak{X}$  expressible as the union of a finite number of elements of  $\mathfrak{X}$  (Ore [1], §2). This union is the required residual.

We list for reference the more important properties of residuation and multiplication (Ward [1], Dilworth [1]):

$$(4.1) \quad a:b \supset a.$$

$$(4.2) \quad a:(a:b) \supset (a, b).$$

$$(4.3) \quad (a:b):c = a:(bc).$$

$$(4.4) \quad [a, b]:b = a:b.$$

$$(4.5) \quad a:(a, b) = a:b.$$

$$(4.51) \quad c:[a, b] \supset (c:a, c:b).$$

$$(4.6) \quad \text{If } a:b = a, \text{ then } a \supset bx \\ \text{implies } a \supset x.$$

$$(4.7) \quad \text{If } r = a:b \text{ and } s = a:r, \text{ then} \\ r = a:s.$$

$$(4.71) \quad a \supset b \text{ implies } ac \supset bc.$$

$$(4.8) \quad [a, b] \supset ab \supset [a, b](a, b).$$

$$(4.81) \quad ab:a \supset b.$$

$$(4.9) \quad a = bc \text{ implies } b \supset a.$$

$$(4.10) \quad (a, b) = i \text{ implies } ab = [a, b] \\ \text{and } (a, bc) = (a, c).$$

$$(4.11) \quad (a, b):c \supset (a:c, b:c).$$

(4.12) In any chain of powers  $a, a^2, a^3, \dots$  either all elements are distinct or all are equal from a certain point on.

\* This axiom is equivalent to the ascending chain condition, as may be seen on taking  $a=b=i$ . We state it in this manner to emphasize the analogy with R 8 which is not equivalent to the descending chain condition.

5. We shall now exhibit a remarkable reciprocity between the operations of residuation and multiplication.

**THEOREM 5.1.** *If a residuation  $x:y$  exists in  $\mathfrak{S}$  satisfying conditions R 1–R 6, and if either of the conditions R 7 or R 8 below holds, then a multiplication  $x \cdot y$  exists in  $\mathfrak{S}$  satisfying M 1–M 6.*

**R 7.**  *$\mathfrak{S}$  is completely closed with respect to cross-cut, and if  $c$  is the cross-cut of a set  $\mathfrak{X}$ , then the cross-cut of the set of all  $a:x$ , where  $a \in \mathfrak{S}$ ,  $x \in \mathfrak{X}$ , equals  $a:c$ .*

**R 8.** *For any two elements  $a, b$  of  $\mathfrak{S}$  the descending chain condition holds in the set  $\mathcal{Y}$  of all elements  $y$  such that  $y:a \supset b$ .*

R 8 is satisfied in many important instances where R 7 does not hold and where the descending chain condition does not hold; for example, in polynomial ideal theory and the classical ideal theory of algebraic rings.

The proof is as follows. Define the "product"  $a \cdot b$  of any two elements  $a$  and  $b$  of  $\mathfrak{S}$ :

**DEFINITION 5.1.** (i)  $a \cdot b:a \supset b$ ; (ii) if  $y:a \supset b$ , then  $y \supset a \cdot b$ .

*Postulate M 1 is satisfied.* For the set  $\mathcal{Y}$  of all  $y$  such that  $y:a \supset b$  is non-empty, since it includes  $b$  by (4.1). If R 7 holds,  $\mathcal{Y}$  has a cross-cut  $p=a \cdot b$  satisfying Definition 5.1, (ii), and the cross-cut  $[\mathcal{Y}:a]$  equals  $p:a$ . Definition 5.1, (i) is therefore satisfied with  $p=a \cdot b$  by the definition of cross-cut.

If R 8 holds, then  $\mathcal{Y}$  again has a cross-cut  $p$  representable as the cross-cut of a finite number of  $y$ ,  $p=[y_1, \dots, y_k]$ . Thus Definition 5.1, (ii) is satisfied, and Definition 5.1, (i) is satisfied by R 5.

*Postulate M 2 is satisfied.* For by R 3,  $a=b$  implies  $a \supset b$  implies  $a \cdot c:b \supset a \cdot c:a$ . Hence by Definition 5.1, (i),  $a \cdot c:b \supset c$  so that by Definition 5.1, (ii),  $a \cdot c \supset b \cdot c$ . Similarly  $a=b$  implies  $b \cdot c \supset a \cdot c$ , so that M 2 follows.

*Postulate M 3 is satisfied.* For  $b \cdot a$  exists, and by Definition 5.1, if  $y:b \supset a$ , then  $y \supset b \cdot a$ . Now by R 4, Definition 5.1, (i) and R 2,  $(a \cdot b:b):a=(a \cdot b:a):b=i$ . Hence by R 2,  $a \cdot b:b \supset a$ . Hence  $a \cdot b \supset b \cdot a$ . Similarly,  $b \cdot a \supset a \cdot b$ ,  $a \cdot b=b \cdot a$ . Condition R 4 is thus seen to insure that multiplication is commutative.

*Postulate M 4 is satisfied.* For by Definition 5.1, (i),  $\{a \cdot (c \cdot b)\}:a \supset c \cdot b$ . Hence  $\{\{a \cdot (c \cdot b)\}:a\}:c \supset c \cdot b:c$  by R 3. But  $c \cdot b:c \supset b$  by Definition 5.1, (i). Therefore  $\{\{a \cdot (c \cdot b)\}:a\}:c \supset b$  or by R 4,  $\{\{a \cdot (c \cdot b)\}:c\}:a \supset b$ . Hence  $\{a(cb):c\} \supset ab$  and  $a(cb) \supset c(ab)$  by Definition 5.1, (ii). Interchanging  $a$  and  $c$ ,  $c(ab) \supset a(cb)$ . Hence  $a(cb)=c(ab)$ , or by M 3 and M 2,  $(ab)c=a(bc)$ .

*Postulate M 5 is satisfied.* For by R 2,  $a:a \supset i$ . Hence  $a \supset ai$  by Definition 5.1, (ii). Now  $ia:i \supset a$  by Definition 5.1, (i). But by (4.10) and M 3,  $ia:i=ia=ai$ . Hence  $ai \supset a$ ,  $a=ai$ .

*Postulate M 6 is satisfied.* For since  $(b, c) \supset b$ ,  $a(b, c) \supset ab$  by (4.71). Similarly  $a(b, c) \supset ac$ . Hence  $a(b, c) \supset (ab, ac)$ . Next  $(ab, ac):a \supset ab:a \supset b$  by R 3

and (4.81). Similarly  $(ab, ac):a \supset c$ . Hence  $(ab, ac):a \supset (b, c)$ . Therefore by Definition 5.1, (ii),  $(ab, ac) \supset a(b, c)$  giving M 6. This completes the proof.

DEFINITION 5.2. (i)  $a \supset (a \circ b)b$ ; (ii) if  $a \supset xb$ , then  $a \circ b \supset x$ .

The following theorem further illustrates the reciprocity between multiplication and residuation:

THEOREM 5.2. If  $a \circ b$  is defined as above, where the multiplication  $xy$  is defined by Definition 5.1, then  $a \circ b = a:b$ .

For since  $a:b \supset a:b$ , we have  $a \supset (a:b)b$  by Definition 5.1, (ii). Therefore by Definition 5.2, (ii),  $a \circ b \supset a:b$ . Now  $a \supset (a \circ b)b$  by Definition 5.2, (i). Therefore by R 3,  $a:b \supset \{(a \circ b)b:b\}$ . But by M 3 and Definition 5.1, (ii),  $(a \circ b)b:b \supset a \circ b$ . Hence  $a:b \supset a \circ b$ ,  $a:b = a \circ b$ .

Hereafter when we speak of a "residuated lattice," we shall mean a lattice in which both a residuation and its associated multiplication are defined satisfying M 1-M 6, R 1-R 6 and the conditions of Definitions 5.1 and 5.2.

6. We may prove by simple examples the following theorem:

THEOREM 6.1. The Dedekind modular condition and the existence of a residual or a multiplication are completely independent properties of a lattice.

It is important to observe that a given lattice may usually be residuated in several different ways. To give a simple example, consider the lattice of four elements  $i > a > b > z$ . The tables for  $x:y$  and  $x \cdot y$  are as follows:

$x:y$	$i$	$a$	$b$	$z$	$x \cdot y$	$i$	$a$	$b$	$z$
$i$	$i$	$i$	$i$	$i$	$i$	$i$	$a$	$b$	$z$
$a$	$a$	$i$	$i$	$i$	$a$	$a$	*	*	$z$
$b$	$b$	*	$i$	$i$	$b$	$b$	*	*	$z$
$z$	$z$	*	*	$i$	$z$	$z$	$z$	$z$	$z$

A brief analysis discloses that the combinations denoted by stars may be determined in six ways so as to satisfy R 1-R 8, M 1-M 8:

	I	II	III	IV	V	VI
$b:a$	$a$	$a$	$b$	$a$	$b$	$b$
$z:a$	$a$	$b$	$b$	$z$	$z$	$z$
$z:b$	$a$	$a$	$a$	$z$	$b$	$z$
$a \cdot a$	$z$	$b$	$a$	$b$	$a$	$a$
$a \cdot b$	$z$	$z$	$z$	$b$	$b$	$b$
$b \cdot b$	$z$	$z$	$z$	$b$	$z$	$b$

Cases II and VI are illustrated in the lattice of the ring of integers modulo 8. Here  $i$  is the set of residue classes  $\{1, 3, 5, 7\}$ ,  $a$  is  $\{2, 6\}$ ,  $b$  is  $\{4\}$ , and  $z$  is  $\{8\}$ . Case II ensues on taking for  $x \cdot y$  multiplication modulo 8, and case VI on taking for  $x \cdot y$  the L.C.M. operation.

The only other lattice of order four is  $i, a, b, z$  with  $(a, b) = i$ ,  $[a, b] = z$ . This lattice may be residuated in only one way, an illustration of a general theorem on the residuation of Boolean algebras which we prove later.

### III. CONDITIONS FOR RESIDUATION

7. In this division of the paper we shall give various sufficient conditions and necessary conditions for the existence of a residuation in a given lattice.

**THEOREM 7.1.** *A necessary condition that a lattice  $\mathfrak{S}$  can be residuated is that any co-prime set of elements of  $\mathfrak{S}$ ,  $a_1, a_2, \dots, a_r$ , generates a Boolean algebra  $\mathfrak{B}$  of order  $2^r$ .*

This condition is not sufficient for a residuation to exist. It is satisfied, for example, in Dedekind's free modular lattice on three elements of order twenty-eight (Dedekind [1], Birkhoff [1], Ore [1]) which we shall prove later cannot be residuated.

Let  $a_1, a_2, \dots, a_r$  be a co-prime set so that

$$(7.1) \quad (a_u, a_v) = i, \quad u, v = 1, \dots, r; u \neq v.$$

The set will remain co-prime if we adjoin  $i$  to it. We shall suppose that this has been done, and for definiteness choose our notation so that  $a_1 = i$ .

Form from the set of  $a$ 's the "ray"  $\Pi$  of  $2^r$  formally distinct cross-cuts:

$$u = [a_{u_1}, a_{u_2}, \dots, a_{u_L}], \quad 1 \leq u_1 < u_2 < \dots < u_L \leq r; 1 \leq L \leq r.$$

We call the  $a_u$  the *constituents* of  $u$ . The ray  $\Pi$  is obviously closed under cross-cut. We shall show that  $\Pi$  is the Boolean algebra required.

**LEMMA 7.1.** *If  $x$  is any element of  $\mathfrak{S}$ , then*

$$(x, [a_u, a_v]) = [(x, a_u), (x, a_v)].$$

This result is trivial if  $u = v$ . But if  $u \neq v$ ,  $(a_u, a_v) = i$ . Hence  $((x, a_u), (x, a_v)) = i$ . Therefore by (4.10) and M 6,

$$\begin{aligned} [(x, a_u), (x, a_v)] &= (x, a_u)(x, a_v) = (x^2, xa_u, a_u x, a_u a_v) \\ &= (x^2, x(a_u, a_v), a_u a_v) = (x, a_u a_v) = (x, [a_u, a_v]) \end{aligned}$$

by M 3 and M 6.

The following two corollaries of this lemma may be proved by induction:

**LEMMA 7.2.** *If  $u = [a_{u_1}, \dots, a_{u_L}]$ , then  $(x, u) = [(x, a_{u_1}), \dots, (x, a_{u_L})]$ .*

LEMMA 7.3. If  $u = [a_{u_1}, \dots, a_{u_L}]$  and  $v = [a_{v_1}, \dots, a_{v_M}]$ , then

$$(u, v) = [(a_{u_1}, a_{v_1}), \dots, (a_{u_L}, a_{v_M})].$$

LEMMA 7.4. If  $x$  is any element of  $\mathfrak{S}$  and if  $(x, b) = (x, c) = i$ , then

$$(x, [b, c]) = [(x, b), (x, c)].$$

It suffices to show that  $(x, [b, c]) = i$ . But  $(x, [b, c]), \supset (x, bc) = (x, bx, bc)$  (by (4.10))  $= (x, b(x, c)) = (x, b) = i$ .

We return to the proof of our theorem. The ray  $\Pi$  is a lattice. For by Lemma 7.3 and (7.1) it is closed under union. The lattice is of order  $2^r$ . It suffices to show that if  $u = v$ , the constituents of  $u$  and  $v$  are identical. But if  $u = v$ ,  $a_u \supset v$ . Hence by Lemma 7.2,

$$a_u = (a_u, v) = [(a_u, a_{v_1}), \dots, (a_u, a_{v_M})].$$

Since  $(a_u, a_v) = a_u$  or  $i$ ,  $a_u$  must be a constituent of  $v$ . Thus every constituent of  $u$  is a constituent of  $v$ . Similarly every constituent of  $v$  is a constituent of  $u$ , so that  $u$  and  $v$  are not formally distinct.

The lattice is distributive. For by Lemma 7.3, if  $w = [a_{w_1}, \dots, a_{w_N}]$ , then

$$\begin{aligned} (w, [u, v]) &= [\dots, (a_w, [u, v]), \dots] \\ &= [\dots, [(a_w, u), (a_w, v)], \dots] \\ &= [[\dots, (a_w, u), \dots], [\dots, (a_w, v), \dots]] \\ &= [(w, u), (w, v)], \end{aligned}$$

by Lemma 7.2.

The lattice is complemented. For we assign to the element  $u$  the complement

$$u' = [a_{u'_L+1}, \dots, a_{u'_r}]$$

where  $u'_{L+1}, \dots, u'_r$  is the selection complementary to  $u_1, \dots, u_L$  from  $1, 2, \dots, r$ . Then  $[u, u'] = [a_1, a_2, \dots, a_r]$ , the null element of the lattice  $\mathfrak{B}$ , and  $(u, u') = i$  by Lemma 7.3. Hence  $\mathfrak{B}$  is a complemented distributive lattice and thus a Boolean algebra.

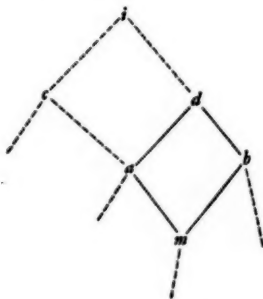
THEOREM 7.2. If  $a_1, \dots, a_r$  is a co-prime set of divisor-free elements of a residuated lattice  $\mathfrak{S}$ , then the Boolean algebra  $\mathfrak{B}$  which they generate is dense over  $\mathfrak{S}$ .

For if  $u$  lies in  $\mathfrak{B}$  and  $x \supset u$ , then  $x = [(x, a_{u_1}), \dots, (x, a_{u_L})]$  by Lemma 7.2. Since  $(x, a_u) = i$  or  $a_u$ , the result follows.

This theorem is quite useful in examining finite lattices to see whether or not they can be residuated. We have also found the following exclusion principle useful in this connection. The proof (which we omit) follows from Lemma 7.4.

**THEOREM 7.3. EXCLUSION PRINCIPLE.** *Let  $a, b, c$ , and  $d$  be any four elements of a residuated lattice  $\mathfrak{S}$  with an ascending chain condition such that  $c \supset a$ ,  $c \neq i$ ;  $(b, c) = i$ ,  $d > a$ ,  $d > b$ . Then if  $m = [b, c]$  we must have  $[a, b] = m$  and  $a > m$ ,  $b > m$ . Furthermore  $a$  and  $b$  are the only elements covered by  $d$  and covering  $m$  in the lattice.*

In schematic form (Klein [1], Birkhoff [2]) the lattice must have the following structure, where the dotted lines indicate that the configuration of the remaining lattice parts is irrelevant.



As a simple application, if the reader will diagram the lattice of order nine on three elements  $b, c$ , and  $f$  where  $c \supset f$  (Dedekind [1]) and take  $a = [c, (f, b)]$ ,  $d = (f, b)$ , he will see that this lattice cannot be residuated.

**THEOREM 7.31.** *The only complemented lattices which can be residuated are Boolean algebras.*

Since by hypothesis the lattice is complemented, it is sufficient to show that it is distributive. We need the following lemma:

**LEMMA 7.5.** *If  $(b, c) = i$  and  $a \supset [b, c]$ , then  $(a:b, a:c) = i$ .*

For we have

$$\begin{aligned} (a:b, a:c) &= (a:b, a:c):(b, c) = [(a:b, a:c):b, (a:b, a:c):c] \\ &\supset [(a:c):b, (a:b):c] = a:cb = a:[c, b] = i. \end{aligned}$$

A complement  $a'$  of  $a$  is defined by the following conditions:

**DEFINITION 7.1.**  $(a, a') = i$ ,  $[a, a'] = z$ , where  $z$  is the null element of  $\mathfrak{S}$ .

Let  $a, b, c$  be any three elements of  $\mathfrak{S}$  and assume that

$$(i) \quad a \supset [b, c].$$

Let  $u = [(a, b), (a, c)]$  and  $v = ([b, c], a')$ . It suffices to show that (i) im-

plies  $u=a$ . We have trivially  $u \supset a$  and  $v \supset a'$ . Hence  $(u, v)=i$  by Definition 7.1 so that  $uv=[u, v]$ . Hence  $b:uv=b:[u, v] \supset (b:u, b:v)$  by (4.51). Now  $b:u \supset b:a$  by R 3 and  $b:v=[b:[b, c], b:a']=b:a'$ . Hence  $b:uv \supset (b:a, b:a')$ . But by Definition 7.1,  $(a, a')=i$  and  $b \supset [a, a']$ . Hence by Lemma 7.5,  $(b:a, b:a')=i$  so that  $b \supset uv$ . Similarly  $c \supset uv$  so that  $[b, c] \supset uv$ , or by (i),  $a \supset uv, a:v \supset u$ . But  $a:v=[a:[b, c], a:a']=a:a'=a$  by (i) and Definition 7.1. Hence  $a \supset u$  so that  $a=u$ .

**COROLLARY.** *The only projective geometries (Birkhoff [3]) which can be residuated are Boolean algebras.*

In case the ascending chain condition holds in  $\mathfrak{S}$ , one can give a much shorter proof by showing that each element may be represented as a cross-cut of a finite number of divisor-free elements and appealing to Theorem 7.1.

**THEOREM 7.4.** *The only multiplication which can be defined over a Boolean algebra is the cross-cut operation.*

In view of our reciprocity theorems it suffices to show that only one residual is definable. One of us has shown elsewhere (Dilworth [1]) that  $a \vee b'$  is a residuation in a Boolean algebra. Suppose that  $a:b$  were another. Then

$$\begin{aligned}(a:b):(a \vee b') &= (a:b):b' = a:bb' = i; \\ (a \vee b'):(a:b) &\supset \{a:(a:b)\} \vee \{b':(a:b)\} \supset b \vee b' = i.\end{aligned}$$

Hence  $a:b = a \vee b'$ .

An interesting consequence of Theorem 7.4 is the following corollary:

**COROLLARY.** *In the ring of integers modulo a square-free integer, the operations of multiplication and L.C.M. are identical.*

8. We consider in this section some sufficient conditions for residuation. We have the following theorem:

**THEOREM 8.1.** *Every lattice in which only one divisor-free element exists can be residuated in at least one way.*

Let  $d$  be the single divisor-free element. We define the residual  $a:b$  by the conditions:

$$(i) \ a:i=a; \quad (ii) \ a:b=i \text{ if } a \supset b; \quad (iii) \ a:b=d \text{ if } a \not\supset b, b \neq i.$$

Then postulates R 1 and R 2 are obviously satisfied.

R 3 is satisfied. For assume  $a \supset b$ . Then  $a:c$  always divides  $b:c$  except possibly when  $b:c=i$ . But then  $b \supset c$ ; so  $a \supset c, a:c=i$ . Similarly  $c:b \supset c:a$ .

R 4 is satisfied. For R 4 obviously holds if  $a, b$ , or  $c$  equals  $i$ . If  $a \supset b, a \neq i$ ,

then  $a:c \supset a \supset b$ ; so  $(a:c):b = (a:b):c = i$ . If  $a:b \supset c$  but  $a \not\supset b$ ,  $b \neq i$ ,  $a \not\supset c$ ,  $c \neq i$ , then  $a:b = a:c = d$ , whence  $(a:b):c = d:c = i = d:b = (a:c):b$ . If  $a \not\supset b$ ,  $a \not\supset c$ ,  $a:b \not\supset c$ ,  $a:c \not\supset b$ , then  $b$  or  $c = i$ .

R 5 is satisfied. For if  $c = i$ , R 5 is trivial. If  $a \supset c$ ,  $b \supset c$ , then  $[a, b] \supset c$  and R 5 obviously holds. If  $a \not\supset c$ ,  $c \neq i$ , then  $[a, b] \not\supset c$  and  $[a, b]:c = d = [a:c, b:c]$ . Hence R 5 holds in general.

In exactly the same way we show that R 6 is satisfied.

F. Klein has shown (Klein [1]) that the modular or distributive properties of a lattice built up of sublattices connected by nodes ("Schnurstellen") depend upon the modular or distributive properties of the sublattices. We prove a similar result for residuation.

**THEOREM 8.2.** *A lattice built up out of a set of residuated lattices connected into a chain by nodes can be residuated.*

It will suffice to prove the theorem for the case of two lattices connected by a node.

Let  $\mathcal{S}$  be composed of two lattices  $\mathcal{S}_1$  and  $\mathcal{S}_2$  connected by a node, so that  $x_1 \in \mathcal{S}_1$  and  $x_2 \in \mathcal{S}_2$  imply  $x_1 \supset x_2$ . Let  $i$  be the unit element of  $\mathcal{S}_1$ . We shall consider the nodal element as belonging to  $\mathcal{S}_1$ , and let  $x_1 y$  denote the residuation in  $\mathcal{S}_1$ ,  $x \circ y$ , the residuation in  $\mathcal{S}_2$  when the nodal element is replaced by  $i$ .

We now define a residual in  $\mathcal{S}$  by the conditions:

$$\begin{aligned} a:b &= a : b \text{ if } a, b \in \mathcal{S}_1, \quad a:b = a \circ b \text{ if } a, b \in \mathcal{S}_2, \\ a:b &= i \text{ if } a \in \mathcal{S}_1, b \in \mathcal{S}_2, \quad a:b = a \text{ if } a \in \mathcal{S}_2, b \in \mathcal{S}_1. \end{aligned}$$

Then postulates R 1, R 2, and R 3 are obviously satisfied.

*Postulate R 4 is satisfied.* For clearly  $a:c \supset a$ . Hence if  $a \supset b$ , then  $(a:b):c = (a:c):b$  by R 3. Also if  $a \in \mathcal{S}_1$ , then  $(a:b):c = (a:c):b$ . Suppose that  $a \in \mathcal{S}_2$ . Then if  $b \in \mathcal{S}_2$ ,  $c \in \mathcal{S}_2$ , we have  $(a:b):c = (a:c):b$ . Similarly if  $b \in \mathcal{S}_1$ ,  $c \in \mathcal{S}_1$ , then  $(a:b):c = (a:c):b$ . Finally if  $b \in \mathcal{S}_1$ ,  $c \in \mathcal{S}_2$ , then  $(a:b):c = a \circ c = (a:c):b$ .

*Postulate R 5 is satisfied.* For R 5 is trivial if  $a, b$ , or  $c = i$ . If  $a \supset b$ , R 5 follows from R 3. If  $a, b \in \mathcal{S}_1$  or  $a, b \in \mathcal{S}_2$ , R 5 holds since it holds in  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

In a similar manner one can show that R 6 is satisfied, and the proof is complete.

By the *direct product* (Birkhoff [4])  $\mathcal{S}$  of the lattices  $\mathcal{S}_1, \dots, \mathcal{S}_n$  we mean the set of vectors  $a = \{a_1, \dots, a_n\}$ , ( $a_i \in \mathcal{S}_i$ ), where the operations are defined by

$$\begin{aligned} [a, b] &= \{[a_1, b_1], \dots, [a_n, b_n]\}, \\ (a, b) &= \{(a_1, b_1), \dots, (a_n, b_n)\}, \end{aligned}$$

and  $a \supset b$  if and only if  $a_i \supset b_i$ , ( $i = 1, \dots, n$ ).

If the  $\mathfrak{S}_i$  are residuated lattices, then  $\mathfrak{S}$  can be residuated, since we may define  $a:b$  to be  $\{a_1:b_1, \dots, a_n:b_n\}$ .

We shall call two sublattices  $\mathfrak{S}_1, \mathfrak{S}_2$  of  $\mathfrak{S}$  *co-prime* if  $a_1 \in \mathfrak{S}_1, a_2 \in \mathfrak{S}_2$  implies that  $(a_1, a_2) = i$ . The sublattices  $\mathfrak{S}_1, \dots, \mathfrak{S}_n$  will be called a *co-prime set* if they are co-prime in pairs.

We note that if  $\mathfrak{S}$  is the direct product of the sublattices  $\mathfrak{S}_1, \dots, \mathfrak{S}_n$  with unit elements, then  $\mathfrak{S}$  contains sublattices  $\mathfrak{S}'_1, \dots, \mathfrak{S}'_n$  simply isomorphic to  $\mathfrak{S}_1, \dots, \mathfrak{S}_n$  and such that  $\mathfrak{S}'_1, \dots, \mathfrak{S}'_n$  is a co-prime set. Birkhoff (Birkhoff [4]) has defined sublattices  $\mathfrak{S}_1, \dots, \mathfrak{S}_n$  to be "strongly" co-prime if each  $\mathfrak{S}_i$  is co-prime to the lattice generated by the remaining lattices. Clearly strong co-primeness implies co-primeness in the ordinary sense. Moreover if  $\mathfrak{S}$  is residuated, Lemma 7.4 shows that co-primeness implies strong co-primeness, so that for residuated lattices the notions are identical. We now prove a converse result.

**THEOREM 8.3.** *Let  $\mathfrak{S}_1, \mathfrak{S}_2, \dots, \mathfrak{S}_n$  be a set of co-prime sublattices of a residuated lattice  $\mathfrak{S}$  such that each element of  $\mathfrak{S}$  can be expressed as a cross-cut of elements of  $\mathfrak{S}_1, \dots, \mathfrak{S}_n$ . Then  $\mathfrak{S}$  is the direct product of  $\mathfrak{S}_1, \dots, \mathfrak{S}_n$ , and each  $\mathfrak{S}_i$  can be residuated.*

Let  $a = [a_1, a_2, \dots, a_n]$ , ( $a_i \in \mathfrak{S}_i$ ),  $b = [b_1, b_2, \dots, b_n]$ , ( $b_i \in \mathfrak{S}_i$ ). Then  $[a, b] = [[a_1, b_1], \dots, [a_n, b_n]]$ . Furthermore

$$(a, b) = [(a_1, b_1), \dots, (a_n, b_n)].$$

For by Lemma 7.2,

$$\begin{aligned} (a, b) &= (a, [b_1, \dots, b_n]) = [(a, b_1), \dots, (a, b_n)] \\ &= [\dots, (a_j, b_k), \dots] = [(a_1, b_1), \dots, (a_n, b_n)], \end{aligned}$$

since  $(a_j, b_k) = i$  if  $j \neq k$ .

**LEMMA 8.1.** *If  $b_1, b_2, \dots, b_n$  are a co-prime set, then*

$$a:[b_1, \dots, b_n] = (\dots ((a:b_1):b_2):b_3) \dots ) : b_n.$$

This result follows by repeated applications of Lemma 7.4, (4.3), and (4.10).

We have now  $a:b = a:[b_1, \dots, b_n] = (\dots ((a:b_1):b_2) \dots ) : b_n$  by Lemma 8.1. But

$$\begin{aligned} a:b_i &= [a_1, \dots, a_n]:b_i = [a_1:b_i, \dots, a_n:b_i] \\ &= [a_1, \dots, a_{i-1}, a_i:b_i, a_{i+1}, \dots, a_n]. \end{aligned}$$

Hence  $a:b = [a_1:b_1, a_2:b_2, \dots, a_n:b_n]$ .

If  $a = b$ , then  $(a_i, [a_1, \dots, a_n]) = (a_i, [b_1, \dots, b_n])$  or  $a_i = (a_i, b_i)$ ,  $a_i \supset b_i$ .

Similarly  $b_i \supset a_i$ . Hence  $\mathfrak{S}$  is simply isomorphic with the direct product of the  $\mathfrak{S}_i$ . We note that if  $x = [x_1, \dots, x_n]$ ,  $(x \supset a_i, a_i \in \mathfrak{S}_i)$ , then

$$x = (x, a_i) = ([x_1, \dots, x_n], a_i) = (x_i, a_i) \in \mathfrak{S}_i.$$

Since  $a_i : b_i \supset a_i$ , we see that  $\mathfrak{S}_i$  is closed under residuation, which completes the proof.

We conclude with a theorem of a more special character.

**THEOREM 8.4.** *The free modular lattice of order twenty-eight on three elements cannot be residuated.*

We shall use Dedekind's original notation for the elements of this lattice in the proof of the theorem (Dedekind [1]). Assume that a residuation  $x:y$  exists. Then  $\delta''''$  is the unit element. Hence  $a_0 : \delta' = b_0 : \delta' = c_0 : \delta' = h \neq \delta''''$ . Now  $a_0 = [a', a''']$ . Hence  $a_0 : a''' = a' : a''' \supset a'$ . But  $a''' \supset \delta'$ . Therefore  $a_0 : \delta' \supset a_0 : a'''$  or  $h \supset a'$ . Similarly,  $h \supset b'$ ,  $h \supset c'$ . Hence  $h \supset (a', b', c')$  or  $h \supset \delta''''$ ,  $h = \delta''''$  giving a contradiction.

It may be observed that the "exclusion principle" of Theorem 7.3 cannot be applied to prove this theorem.

#### IV. NOETHER LATTICES\*

9. Consider any residuated lattice  $\mathfrak{S}$ . An element  $c$  of  $\mathfrak{S}$  is *irreducible* if in every decomposition  $c = [g, f]$  into a cross-cut of two elements of  $\mathfrak{S}$ , either  $g = c$  or  $f = c$ . An element  $p$  is a *prime* if  $p \supset ab$  implies  $p \supset a$  or  $p \supset b$ , and *primary* if  $p \supset ab$ ,  $p \nmid a$  implies  $p \supset b^s$  for some integer  $s$ . The irreducible elements are thus determined by an intrinsic lattice property, while the primes and primary elements depend upon the particular multiplication introduced into the lattice.

We propose here the name "Noether lattice" for any lattice  $\mathfrak{S}$  satisfying the following three conditions:

- N 1. *The lattice  $\mathfrak{S}$  may be residuated.*
- N 2. *The ascending chain condition holds in  $\mathfrak{S}$ .*
- N 3. *Every irreducible element of  $\mathfrak{S}$  is primary.*

By N 1 we mean that  $\mathfrak{S}$  is closed under operations  $x:y$ ,  $xy$  having the properties R 1-R 6, M 1-M 6 and connected by the relationships expressed

\* Our definition differs from that in Ward and Dilworth [1]. We have found that some of the results stated in §4 of this paper are in error. In postulate D 1, the exponent  $r$  must be replaced by 1. The condition  $ab = [a, b]$  on the idempotent elements of a finite modular lattice is consequently necessary for the truth of D 1 but not sufficient. The postulate M 7 is not a sufficient condition for a Noether lattice as stated in the theorem preceding M 7.

in Definitions 4.1, 5.1. By N 2 we mean (Ore [1]) that every chain of lattice elements  $a_1 < a_2 < a_3 < \dots < i$  terminates.

We choose the name in honor of Emmy Noether because the decomposition theorems first proved by her for the ideals of a commutative ring with chain condition all hold. It is to be observed that we do not assume a modular condition.

The proof that the usual decomposition theorems hold may be made by a mere transcription of the proofs given in van der Waerden [1] into lattice language. With each primary  $q$  is associated a prime  $p$  with the properties  $p \supset q$  and  $p \supset b$  implies  $q \supset b^*$ . A cross-cut of primaries is said to be "simple" if no primary in it divides the cross-cut of any of the remaining primaries. *Every element not equal to  $i$  of a Noether lattice may be represented as a simple cross-cut of a finite number of primaries each of which is associated with a different prime. The primes themselves and the total number of primaries are uniquely determined by the element.* We obtain from each such representation a representation as the cross-cut of "isolated components" by grouping together the cross-cuts of primaries whose associated primes divide one another. *The isolated components of an element and the corresponding representation as their cross-cut are unique.*

As was pointed out by Krull (Krull [1]), the decompositions into relatively prime ("teilerfremd") elements depend merely upon N 1 and N 2. From our standpoint, they are simple consequences of Theorem 7.1 and the chain condition.

We may specialize our lattice still more by the following assumption:

N 4. *Every prime of  $\mathfrak{S}$  is divisor-free.*

Then, since we have trivially from N 1 that every divisor-free element is a prime, we easily see that all primaries associated with a given prime form a lattice which we may say "belongs" to this prime.

The lattices belonging to distinct primes have no elements save  $i$  in common. Hence the decomposition theorems in this case are merely an instance of Birkhoff's decomposition of a lattice into direct products relative to cross-cut (Birkhoff [4]).

10. We shall now give some general properties of any Noether lattice.

**THEOREM 10.1.** *If  $a$  and  $b$  are any two elements of a Noether lattice, there exists an exponent  $s$  such that the following condition holds:*

D 1.  $ab \supset [a, b^s]$ .

Let  $ab = [q_1, \dots, q_k]$  be a decomposition of  $ab$  into a cross-cut of primaries. Then for each  $q_i$ ,  $q_i \supset ab$ ; hence either  $q_i \supset a$  or  $q_i \nsubseteq a$ ,  $q_i \supset b^s$ . With

a proper choice of notation, we may assume that  $q_i \supset a$ , ( $i=1, \dots, l$ ),  $q_i \supset b^{s_i}$ , ( $i=l+1, \dots, k$ ). Hence if  $s$  is the largest of the  $s_i$ ,  $[q_1, \dots, q_l] \supset a$ ,  $[q_{l+1}, \dots, q_k] \supset b^s$  giving D 1.

The following three theorems are immediate corollaries:

**THEOREM 10.2.** *If  $b$  is an idempotent element in a Noether lattice, then  $[a, b] = ab$  for any other element  $a$  of the lattice, and  $b[a, c] = [ba, bc]$  for any elements  $a$  and  $c$ .*

**THEOREM 10.3.** *In a Noether lattice, every idempotent element is neutral.\**

**THEOREM 10.4.** *In a Noether lattice, the idempotent elements form a distributive lattice. The product of any two idempotent elements is their cross-cut.*

It is easy to show that this last mentioned property of idempotent elements holds in any lattice in which multiplication is distributive with respect to cross-cut; for if  $a, b$  are idempotent,

$$\begin{aligned} ab \supset [a, b](a, b) &= [a(a, b), b(a, b)] \\ &= [(a^2, ab), (ba, b^2)] = [(a, ab), (b, ab)] = [a, b]. \end{aligned}$$

The following lattice of order six illustrates how the definition of a Noether lattice depends upon the type of multiplication introduced. The elements are  $i, j, a, b, k$ , and  $z$  with the coverings  $i > j, j > a, j > b; a > k, b > k; k > z$ . The lattice is distributive and hence a Noether lattice if multiplication is identified with cross-cut (see §11). Define an operation  $xy$  by  $ix = xi = x; zx = xz = z; j^2 = j, a^2 = a, b^2 = b, k^2 = z; ja = aj = a; jb = bj = b; jk = kj = z; ab = ba = ak = ka = bk = kb = z$ . It may be shown that  $xy$  is a multiplication satisfying M 1-M 8. But D 1 does not hold; for  $ab = z$  and  $[a, b] = k$ , while  $a$  and  $b$  are idempotent. Hence N 3 is false by Theorem 10.1.

11. We shall next give some sufficient conditions that a lattice be a Noether lattice.

**THEOREM 11.1.** *Let  $\mathfrak{S}$  be a residuated lattice with ascending chain condition. Then sufficient conditions that  $\mathfrak{S}$  be a Noether lattice are as follows:*

D 1.  $ab \supset [a, b^*]$ .

D 2.  $\mathfrak{S}$  is modular.

It suffices to show that N 3 holds. Let  $m$  be irreducible,  $m \supset ab$ ,  $m \nsubseteq a$ . Then if  $d = (a, m)$ ,  $d \supset m \supset db$ . Now by D 1,  $db \supset [d, b^*]$  for some  $s$ . Hence  $d \supset m \supset [d, b^*]$ . Therefore by D 2,  $m = [(m, d), (m, b^*)]$ . Since  $m$  is irreducible and  $(m, d) = (m, a) \neq m$ ,  $(m, b^*) = m$ . Hence  $m \supset b^*$  and  $m$  is primary.

\* Following Ore [1], we call an element  $n$  of a lattice "neutral" if  $[n, (b, c)] = ([n, b], [n, c])$  for every pair of elements  $b, c$  of the lattice.

**COROLLARY.** *Every distributive lattice in which the ascending chain condition holds is a Noether lattice for a suitably defined multiplication.*

We take for the multiplication the cross-cut operation. Then M 1-M 6 and M 8 all hold; so  $\mathfrak{S}$  may be residuated. Since  $\mathfrak{S}$  is distributive, it is modular. Thus N 1, N 2, and D 2 hold. But D 1 is trivially true. The result now follows from the previous theorem.

We shall next give some conditions enabling us to view the ideal theory of commutative rings from a lattice-theoretic standpoint. It is first necessary to introduce a new concept. Let  $\mathfrak{S}$  be a residuated lattice.

**DEFINITION 11.1.** *An element  $q$  of  $\mathfrak{S}$  is principal if  $q \supset b$  implies that there exists an element  $c$  such that  $qc = b$ .*

Neither  $c$  nor  $b$  need be principal.

Suppose that  $a$  is principal,  $a \supset b$ . The set  $\mathfrak{J}$  of elements  $z$  such that  $az = b$  is closed with respect to union. If either postulate M 7 or M 8 holds, the union  $b/a$  of  $\mathfrak{J}$  has the properties stated in the following definition:

**DEFINITION 11.2.**  *$a \cdot (b/a) = b$ ; if  $ax = b$  then  $b/a \supset x$ .*

We call  $b/a$  the *quotient* of  $b$  by  $a$ . It is easily shown (Ward [1]) that if  $a$  is principal and  $a \supset b$ , then the quotient  $b/a$  equals the residual  $b:a$  of  $a$  with respect to  $b$ .

As a simple consequence, we have the following lemma:

**LEMMA 11.1.** *If  $a$  is principal and if  $a \supset b$ , then  $b = (b:a)a$ .*

We may observe that M 8 always holds if the ascending chain condition holds. Hence Lemma 11.1 is true for all principal elements of a residuated lattice with ascending chain condition. We shall now prove the following fundamental theorem:

**THEOREM 11.2.** *Let  $\mathfrak{S}$  be a lattice in which the following conditions hold:*

- N 1. *The lattice  $\mathfrak{S}$  may be residuated.*
- N 2. *The ascending chain condition holds in  $\mathfrak{S}$ .*
- D 2.  *$\mathfrak{S}$  is modular.*
- D 3. *Every element of  $\mathfrak{S}$  is the union of a finite number of principal elements.*
- D 4. *The principal elements of  $\mathfrak{S}$  are closed under multiplication.*

*Then  $\mathfrak{S}$  is a Noether lattice.*

The instance of ideal theory is obtained by identifying the principal elements of the lattice with the principal ideals or the corresponding ring ele-

ments. D 3 is then the basis theorem, and D 4 the closure property of ring multiplication.

It suffices to show that every irreducible element is primary, or inversely that every non-primary element is reducible. Let  $m$  be non-primary. Then there exist elements  $a$  and  $b$  of the lattice such that

$$(11.1) \quad m \supset ab, m \not\supset a, m \not\supset b^r \text{ any } r.$$

We shall show that  $m$  is reducible. By D 3,  $b = (b_1, b_2, \dots, b_k)$  where the  $b_i$  are principal. Then  $m \supset ab_i$ . For at least one  $b_i$ ,  $m \not\supset b_i^r$  for any  $r$ . For otherwise, for each  $b_i$  there exists an exponent  $r_i$  such that  $m \supset b_i^{r_i}$ . Then if  $r > r_1 + r_2 + \dots + r_k - 1$ , we have  $m \supset b^r$  contrary to hypothesis. Therefore, we may assume that  $b$  in (11.1) is principal.

By N 2, the chain  $m:b, m:b^2, \dots, m:b^k, \dots$  terminates so that  $m:b^k = m:b^{k+1}$  for some fixed  $k$ . Consider the cross-cut  $c = [(m, a), (m, b^k)]$ . We have trivially  $c \supset m$ . Now  $(m, b^k) \supset c \supset m$ . Hence by D 2 (Ore [1]),

$$(11.2) \quad c = ([c, m], [c, b^k]).$$

Now  $m \supset [c, m]$ . We shall show next that  $m \supset [c, b^k]$ . By D 4,  $b^k$  is principal, and  $b^k \supset [c, b^k]$ . Hence by Lemma 11.1,  $[c, b^k] = \{[c, b^k]: b^k \mid [c, b^k] = (c:b^k)b^k$ . Also since  $(m, a) \supset c$ ,  $b(m, a) \supset bc$ . But  $b(m, a) = (bm, ba) \subset m$  by (11.1). Hence  $m \supset bc \supset b[c, b^k]$  by (11.2). That is,  $m \supset b\{(c:b^k)b^k\}$  or  $m:b^{k+1} \supset c:b^k$ . But  $m:b^{k+1} = m:b^k$ . Hence  $m:b^k \supset c:b^k$  or  $m \supset (c:b^k)b^k$ ,  $m \supset [c, b^k]$ . It follows therefore that  $m \supset c$ . Hence  $m = c$  or  $m = [(m, a), (m, b^k)]$ . But  $m \not\supset a$ ,  $m \not\supset b^k$ . Hence  $(m, a) \neq m$ ,  $(m, b^k) \neq m$ , and  $m$  is reducible. This completes the proof.

12. To show the significance of the hypotheses of Theorems 11.1 and 11.2, we shall exhibit various lattices in which not all the hypotheses are satisfied.

We first consider the following lattice  $\mathfrak{B}_1$ ; and we define a multiplication  $xy$  over  $\mathfrak{B}_1$  by the following table:

$x \backslash y$	$i$	$j$	$a$	$b$	$m$	$z$
$i$	$i$	$j$	$a$	$b$	$m$	$z$
$j$	$j$	$a$	$a$	$z$	$z$	$z$
$a$	$a$	$a$	$a$	$z$	$z$	$z$
$b$	$b$	$z$	$z$	$z$	$z$	$z$
$m$	$m$	$z$	$z$	$z$	$z$	$z$
$z$	$z$	$z$	$z$	$z$	$z$	$z$

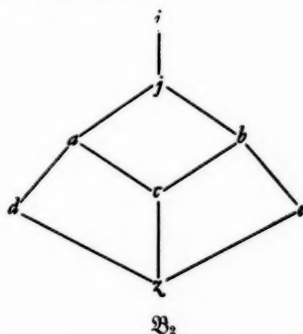
The reader may verify that M 1-M 8 are satisfied. Thus  $\mathfrak{B}_1$  is a residuated lattice in which the ascending chain condition holds.  $\mathfrak{B}_1$  is obviously non-modular. Now it is easily verified that D 1 holds in the lattice:  $xy \supset [x, y^2]$ ,

for every  $x, y$  of the lattice. Nevertheless, *not every irreducible element is primary*. For consider the irreducible  $m$ . We have  $m \supset ab$  and  $m \not\supset b$ . But since  $a$  is idempotent,  $m \not\supset a^r$  for any  $r$ .



Next, consider the lattice  $\mathfrak{B}_2$ .

We assign the residuation  $x:y$  to  $\mathfrak{B}_2$  described in Theorem 8.1. The associated multiplication given by Definition 5.1 is then as follows:  $xy=y$ , if  $x=i$ ;  $xy=x$  if  $y=i$ ;  $xy=x$  otherwise.



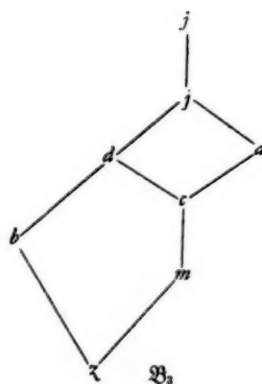
This lattice is non-modular, as it contains the non-modular sublattice  $j, a, d, e, z$ . The irreducible elements in it are  $j, a, b, d, e$ , and these are all primary since  $x \supset y^2$  for any  $y \neq i$  and any  $x$ . Furthermore, the elements  $i, c, d, e$ , and  $z$  are principal and  $a = (d, c)$ ,  $b = (c, e)$ ,  $j = (a, b)$ . Finally the principal

elements are closed with respect to multiplication. Thus in this lattice, all hypotheses of Theorem 11.2 hold save modularity; and yet the lattice is a Noether lattice.

Our last example is one in which all the hypotheses of Theorem 11.2 hold save modularity and the lattice is *not* a Noether lattice. We define a multiplication over  $\mathfrak{B}_3$  by the following table:

$x \backslash y$	$i$	$j$	$a$	$b$	$c$	$d$	$m$	$z$
$i$	$i$	$j$	$a$	$b$	$c$	$d$	$m$	$z$
$j$	$j$	$d$	$c$	$b$	$m$	$d$	$m$	$z$
$a$	$a$	$c$	$c$	$z$	$m$	$m$	$m$	$z$
$b$	$b$	$b$	$z$	$b$	$z$	$b$	$z$	$z$
$c$	$c$	$m$	$m$	$z$	$m$	$m$	$m$	$z$
$d$	$d$	$d$	$m$	$b$	$m$	$d$	$m$	$z$
$m$	$m$	$m$	$m$	$z$	$m$	$m$	$m$	$z$
$z$	$z$	$z$	$z$	$z$	$z$	$z$	$z$	$z$

Then it may be verified that the multiplication satisfies M 1–M 8, and that the elements  $i, a, b, c, m, z$  are principal and closed under multiplication.



Since  $d = (b, m)$ ,  $j = (a, d)$ , every element is the union of a finite number of principal elements. The lattice is evidently non-modular. It is not a Noether lattice. For consider the irreducible element  $m$ . Then  $m \supset ab$ ,  $m \not\supset a$ , and  $m \not\supset b$  for any  $s$ , since  $b$  is idempotent.

## V. CONDITIONS FOR DISTRIBUTIVITY

13. We shall conclude by answering some of the questions raised in Ward [4] as to the import of certain auxiliary conditions in a residuated lattice. We consider a residuated lattice in which one or more of the following conditions hold:

R 9.  $(a:b, b:a) = i$ . R 10.  $a:[b, c] = (a:b, a:c)$ . R 11.  $(b, c):a = (b:a, c:a)$ .

THEOREM 13.1. R 9, R 10, R 11 are equivalent and imply distributivity.

R 9 implies R 11. For

$$\begin{aligned}(b:a, c:a): \{ (b, c):a \} &\supset ((b:a): \{ (b, c):a \}, (c:a): \{ (b, c):a \}) \\ &= ((b: \{ (b, c):a \}):a, (c: \{ (b, c):a \}):a).\end{aligned}$$

But  $(b: \{ (b, c):a \}):a \supset b:c$  since

$$\begin{aligned}\{ (b: \{ (b, c):a \}):a \}: (b:c) &= (\{ b: (b:c) \}: \{ (b, c):a \}):a \\ &\supset ((b, c): \{ (b, c):a \}):a \supset a:a = i.\end{aligned}$$

Similarly  $(c: \{ (b, c):a \}):c \supset c:b$ . Hence  $(b:a, c:a): \{ (b, c):a \} \supset (b:c, c:b) \supset i$  by R 9. Thus  $(b:a, c:a) \supset (b, c):a$ . But  $(b, c):a \supset (b:a, c:a)$  trivially.

R 11 implies R 10. For by R 11,

$$\begin{aligned}(a:b, a:c): \{ a:[b, c] \} &= ((a:b): \{ a:[b, c] \}, (a:c): \{ a:[b, c] \}) \\ &= ((a: \{ a:[b, c] \}):b, (a: \{ a:[b, c] \}):c) \\ &\supset ([b, c]:b, [b, c]:c) \\ &= (c:b, b:c) = (c:(b, c), b:(b, c)) = (c, b):(b, c) = i\end{aligned}$$

by R 11. Hence  $(a:b, a:c) \supset a:[b, c]$ . But  $a:[b, c] \supset (a:b, a:c)$  trivially.

R 10 implies R 9. For  $(a:b, b:a) = ([a, b]:b, [a, b]:a) = [a, b]:[a, b] = i$  by condition R 10.

R 10 implies distributivity. For let  $a \supset [b, c]$ . Then

$$a: [(a, b), (a, c)] = (a:(a, b), a:(a, c)) = (a:b, a:c) = a:[b, c] = i.$$

Hence  $a \supset [(a, b), (a, c)]$  and  $[(a, b), (a, c)] \supset a$  trivially. Therefore  $a = [(a, b), (a, c)]$ .

THEOREM 13.2. If every element of a residuated lattice is principal, then the lattice is distributive.

Let  $(b, c) \supset a$ . We have  $a \supset ([a, b], [a, c])$ . Hence

$$\begin{aligned}a:(b, c) \supset ([a, b], [a, c]):(b, c) &= [([a, b], [a, c]):b, ([a, b], [a, c]):c] \\ &\supset [[a, b]:b, [a, c]:c] = [a:b, a:c] = a:(b, c).\end{aligned}$$

Thus  $a:(b, c) = ([a, b], [a, c]):(b, c)$ . But  $(b, c) \supset a \supset ([a, b], [a, c])$ . Hence

$$a = (a:(b, c))(b, c) = \{([a, b], [a, c]):(b, c)\}(b, c) = ([a, b], [a, c])$$

by Lemma 13.1.

**THEOREM 13.3.** *A sufficient condition that a residuated lattice with ascending chain condition be a Noether lattice is that every element in it be principal.*

For by Theorem 13.2, the lattice is distributive and hence modular; so all the hypotheses of Theorem 11.2 are satisfied.

#### REFERENCES

##### GARRETT BIRKHOFF

1. *On the combination of sub-algebras*, Proceedings of the Cambridge Philosophical Society, vol. 39 (1933), pp. 441-464.
2. *On the structure of abstract algebras*, *ibid.*, vol. 41 (1935), pp. 433-454.
3. *Combinatorial relations in projective geometries*, Annals of Mathematics, (2), vol. 36 (1935), pp. 743-748.
4. *On the lattice theory of ideals*, Bulletin of the American Mathematical Society, vol. 40 (1934), pp. 613-619.

##### R. DEDEKIND

1. *Ueber die von drei Moduln erzeugte Dualgruppe*, Gesammelte mathematische Werke, vol. 2, 1931, paper 30, pp. 236-271.

##### R. P. DILWORTH

1. *Abstract residuation over lattices*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 262-268.

##### FRITZ KLEIN

1. *Dedekindsche und distributive Verbände*, Mathematische Zeitschrift, vol. 41, pp. 261-280.

##### G. KÖTHE

1. *Die Theorie der Verbände . . .*, Jahresbericht der deutschen Mathematiker-Vereinigung, vol. 47 (1937), pp. 125-142.

##### W. KRULL

1. *Axiomatische Begründung der allgemeinen Idealtheorie*, Sitzungsberichte der physikalisch-medizinischen Societät zu Erlangen, vol. 56 (1924), pp. 47-63.

##### O. ORE

1. *On the foundation of abstract algebra, I*, Annals of Mathematics, (2), vol. 36 (1935), pp. 406-437.

##### B. L. VAN DER WAERDEN

1. *Moderne Algebra*, vol. 2, Berlin, 1931.

##### M. WARD AND R. P. DILWORTH

1. *Residuated lattices*, Proceedings of the National Academy of Sciences, vol. 24 (1938), pp. 162-164.

##### M. WARD

1. *Residuation in structures over which a multiplication is defined*, Duke Mathematical Journal, vol. 3 (1937), pp. 627-636.
2. *Structure residuation*, Annals of Mathematics, (2), vol. 39 (1938), pp. 558-568.

CALIFORNIA INSTITUTE OF TECHNOLOGY,  
PASADENA, CALIF.

# KAKEYA'S PROBLEM ON THE ZEROS OF THE DERIVATIVE OF A POLYNOMIAL\*

BY

MORRIS MARDEN†

1. **Introduction.** If all  $n$  zeros of a polynomial  $f(z)$  of degree  $n$  lie in or on a circle  $K$  of radius  $R$ , then, according to the well known theorem of Gauss and Lucas,‡ all  $n-1$  zeros of its derivative  $f'(z)$  also lie in or on  $K$ . If only two zeros of  $f(z)$  lie in or on  $K$ , then, according to a theorem stated by Alexander and proved byakeya and Szegő,§ at least one zero of  $f'(z)$  lies in or on the concentric circle of radius  $R \csc(\pi/n)$ . If all but one of the zeros of  $f(z)$  lie in or on  $K$ , then, according to a theorem due to Biernacki,|| at most one zero of  $f'(z)$  lies outside of the concentric circle of radius  $R(1+1/n)^{1/2}$ . In general, according to a theorem stated byakeya,§ if  $p$  zeros of a polynomial  $f(z)$  of degree  $n$ , ( $2 \leq p \leq n$ ), lie in or on a circle of radius  $R$ , then at least  $p-1$  zeros of its derivative lie in or on a concentric circle of radius  $R\rho(n, p)$ .

The existence of a function  $\rho(n, p)$  was proved byakeya§ in the general case. The actual computation of  $\rho(n, p)$  seems, however, to have been made so far only in the three cases mentioned above; namely,

$$\rho(n, n) = 1, \quad \rho(n, 2) \leq \csc \pi/n, \quad \rho(n, n-1) \leq (1 + 1/n)^{1/2}.$$

Although in the present note the minimum value of  $\rho(n, p)$  in the general case will not be determined, two inequalities for  $\rho(n, p)$  will be established. First, for all  $n$  and  $p$ , ( $2 \leq p \leq n$ ),

$$(1) \quad \rho(n, p) \leq \csc \frac{\pi}{2(n-p+1)},$$

and, secondly, for at least  $p$  an even integer,¶

\* Presented to the Society, September 1, 1936; received by the editors December 24, 1937 and, in revised form, May 19, 1938.

† The results of this paper were obtained under a grant for summer research from the Wisconsin Alumni Research Foundation.

‡ See references in M. Marden, *American Mathematical Monthly*, vol. 42 (1935), pp. 278-279.

§ J. W. Alexander, *Annals of Mathematics*, (2), vol. 17 (1915), p. 18; S.akeya, *Tôhoku Mathematical Journal*, vol. 11 (1917), pp. 5-16, especially p. 9; G. Szegő, *Mathematische Zeitschrift*, vol. 13 (1932), pp. 28-55.

|| M. Biernacki, *Bulletin de l'Académie Polonaise*, 1927, pp. 660-675; See also J. L. Walsh, *these Transactions*, vol. 24 (1922), p. 37.

¶ It is to be noted that for  $p=n$ ,  $\csc \pi/[2(n-p+1)] = 1 = (2-p/n)^{1/2}$ ; and, for  $p < n$ ,  $\csc \pi/[2(n-p+1)] \geq 2^{1/2} > (2-p/n)^{1/2}$ .

$$(2) \quad \rho(n, p) \geq (2 - p/n)^{1/2}.$$

The second inequality may be proved simply by exhibiting a polynomial of degree  $n$  which has  $p=2m$  zeros in or on the unit circle and of which the derivative has at least  $p-1$  zeros in or on the circle  $|z| = (2 - p/n)^{1/2}$ . Such a polynomial is

$$f(z) = \left[ z^2 - 2z \left( \frac{n}{2n-p} \right)^{1/2} + 1 \right]^{p/2} \left[ z - \frac{1}{p} (n(2n-p))^{1/2} \right]^{n-p};$$

for, it has zeros of multiplicity  $p/2$  on the unit circle at the points

$$z = \left( \frac{n}{2n-p} \right)^{1/2} \pm i \left( \frac{n-p}{2n-p} \right)^{1/2},$$

and its derivative has zeros of multiplicity  $(p-2)/2$  at these points and a double zero at the point  $z = (2 - p/n)^{1/2}$ .

The proof of the first inequality, however, will require the establishment of an identity (apparently new) relating any  $p$  zeros of a polynomial

$$f(z) = (z - \alpha_1)(z - \alpha_2) \cdots (z - \alpha_n)$$

with any  $(n-p+1)$  zeros of its derivative which are distinct from the  $p$  given zeros of  $f(z)$ . The identity is a generalization of the well known formula

$$\sum_{j=1}^n \frac{1}{\beta - \alpha_j} = 0$$

relating the  $n$  zeros of  $f(z)$  with any one zero  $\beta$  of  $f'(z)$  which is not a zero of  $f(z)$ .

The identity in question is derived in §2 and applied to the proof of inequality (1) in §3. In §4, the relation of this inequality to one given by Fekete is discussed. Finally, in §5, the inequality is used to obtain a sufficient condition for a polynomial to be at most  $p$ -valent in a given circle or other convex region.

**2. An identity.** The identity mentioned above is described in the following theorem:

**THEOREM 1.** *If the  $n+1$  complex numbers*

$$\alpha_1, \alpha_2, \dots, \alpha_p; \beta_1, \beta_2, \dots, \beta_q, \quad 2 \leq p \leq n, q = n - p + 1,$$

*are distinct and if all of the  $\alpha_j$  are zeros of a polynomial  $f(z)$  of degree  $n$  and all of the  $\beta_k$  are zeros of its derivative  $f'(z)$ , then*

$$(3) \quad \sum \frac{D_{j_1 j_2 \dots j_q}}{(\beta_1 - \alpha_{j_1})(\beta_2 - \alpha_{j_2}) \cdots (\beta_q - \alpha_{j_q})} = 0,$$

where  $j_1, j_2, \dots, j_q$  run independently from 1 to  $p$ , where

$$D_{j_1 j_2 \dots j_q} = \prod_{m=1}^p (\delta_{mj_1} + \delta_{mj_2} + \dots + \delta_{mj_q})!,$$

and where  $\delta_{mj} = 1$  or 0 according as  $j = m$  or  $j \neq m$ .

To prove Theorem 1, we shall let

$$P(z) = (z - \alpha_1)(z - \alpha_2) \dots (z - \alpha_p).$$

Then there exist  $q$  constants  $a_0, a_1, \dots, a_{q-1}$ , not all zero, such that

$$f(z) = (a_0 + a_1 z + \dots + a_{q-1} z^{q-1})P(z).$$

These constants satisfy the system of  $q$  homogeneous linear equations

$$f'(\beta_j) = a_0 \frac{d}{d\beta_j} P(\beta_j) + a_1 \frac{d}{d\beta_j} [\beta_j P(\beta_j)] + \dots + a_{q-1} \frac{d}{d\beta_j} [\beta_j^{q-1} P(\beta_j)] = 0,$$

$$j = 1, 2, \dots, q,$$

of which system the determinant

$$(4.1) \quad \Delta(\beta_1, \beta_2, \dots, \beta_q) = \begin{vmatrix} \frac{d}{d\beta_1} P(\beta_1) & \frac{d}{d\beta_1} [\beta_1 P(\beta_1)] & \dots & \frac{d}{d\beta_1} [\beta_1^{q-1} P(\beta_1)] \\ \frac{d}{d\beta_2} P(\beta_2) & \frac{d}{d\beta_2} [\beta_2 P(\beta_2)] & \dots & \frac{d}{d\beta_2} [\beta_2^{q-1} P(\beta_2)] \\ \dots & \dots & \dots & \dots \\ \frac{d}{d\beta_q} P(\beta_q) & \frac{d}{d\beta_q} [\beta_q P(\beta_q)] & \dots & \frac{d}{d\beta_q} [\beta_q^{q-1} P(\beta_q)] \end{vmatrix}$$

must therefore vanish.

Defining  $V(\beta_1, \beta_2, \dots, \beta_q)$  as the Vandermonde determinant

$$(4.2) \quad V(\beta_1, \beta_2, \dots, \beta_q) = \begin{vmatrix} 1 & \beta_1 & \beta_1^2 & \dots & \beta_1^{q-1} \\ 1 & \beta_2 & \beta_2^2 & \dots & \beta_2^{q-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \beta_q & \beta_q^2 & \dots & \beta_q^{q-1} \end{vmatrix} = \prod_{j=1}^q \prod_{k=j+1}^q (\beta_k - \beta_j),$$

we may write

$$\Delta(\beta_1, \beta_2, \dots, \beta_q) = \frac{\partial^q}{\partial \beta_1 \partial \beta_2 \dots \partial \beta_q} [P(\beta_1)P(\beta_2) \dots P(\beta_q)V(\beta_1, \beta_2, \dots, \beta_q)],$$

and hence

$$\frac{\partial^{k_1+k_2+\dots+k_q-q}\Delta}{\partial\beta_1^{k_1}\partial\beta_2^{k_2}\dots\partial\beta_q^{k_q}} = \frac{\partial^{k_1+k_2+\dots+k_q}}{\partial\beta_1^{k_1}\partial\beta_2^{k_2}\dots\partial\beta_q^{k_q}} [P(\beta_1)P(\beta_2)\dots P(\beta_q)V].$$

The right-hand side of this equation may be evaluated by Leibniz' rule for differentiating a product, as follows. First,

$$\frac{\partial^{k_1}}{\partial\beta_1^{k_1}} [P(\beta_1)V] = \sum_{j_1=0}^{k_1} C_{k_1,j_1} P^{(k_1-j_1)}(\beta_1) \frac{\partial^{j_1}V}{\partial\beta_1^{j_1}},$$

where  $C_{k_1,j_1} = k_1!/j_1!(k_1-j_1)!$  and  $C_{k_1,0} = 1$ . If, now, we assume that for some fixed value of  $m$ , ( $1 \leq m \leq q$ ),

$$\begin{aligned} \frac{\partial^{k_1+k_2+\dots+k_m}}{\partial\beta_1^{k_1}\partial\beta_2^{k_2}\dots\partial\beta_m^{k_m}} [P(\beta_1)P(\beta_2)\dots P(\beta_m)V] \\ = \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \dots \sum_{j_m=0}^{k_m} \left\{ \prod_{i=1}^m C_{k_i,j_i} P^{(k_i-j_i)}(\beta_i) \frac{\partial^{j_i}}{\partial\beta_i^{j_i}} \right\} V, \end{aligned}$$

then

$$\begin{aligned} \frac{\partial^{k_1+k_2+\dots+k_{m+1}}}{\partial\beta_1^{k_1}\partial\beta_2^{k_2}\dots\partial\beta_{m+1}^{k_{m+1}}} [P(\beta_1)P(\beta_2)\dots P(\beta_{m+1})V] \\ = \sum_{j_1=0}^{k_1} \dots \sum_{j_m=0}^{k_m} \left\{ \prod_{i=1}^m C_{k_i,j_i} P^{(k_i-j_i)}(\beta_i) \frac{\partial^{j_i}}{\partial\beta_i^{j_i}} \right\} \frac{\partial^{k_{m+1}}}{\partial\beta_{m+1}^{k_{m+1}}} P(\beta_{m+1})V \\ = \sum_{j_1=0}^{k_1} \dots \sum_{j_m=0}^{k_m} \left\{ \prod_{i=1}^m C_{k_i,j_i} P^{(k_i-j_i)}(\beta_i) \frac{\partial^{j_i}}{\partial\beta_i^{j_i}} \right\} \\ \cdot \sum_{j_{m+1}=0}^{k_{m+1}} C_{k_{m+1},j_{m+1}} P^{(k_{m+1}-j_{m+1})}(\beta_{m+1}) \frac{\partial^{j_{m+1}}}{\partial\beta_{m+1}^{j_{m+1}}} V \\ = \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \dots \sum_{j_{m+1}=0}^{k_{m+1}} \left\{ \prod_{i=1}^{m+1} C_{k_i,j_i} P^{(k_i-j_i)}(\beta_i) \frac{\partial^{j_i}}{\partial\beta_i^{j_i}} \right\} V. \end{aligned}$$

We may conclude by mathematical induction, therefore, that

$$\begin{aligned} \frac{\partial^{k_1+k_2+\dots+k_q-q}\Delta}{\partial\beta_1^{k_1-1}\partial\beta_2^{k_2-1}\dots\partial\beta_q^{k_q-1}} \\ (5.1) \quad = \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \dots \sum_{j_q=0}^{k_q} \left\{ \prod_{i=1}^q C_{k_i,j_i} P^{(k_i-j_i)}(\beta_i) \frac{\partial^{j_i}}{\partial\beta_i^{j_i}} \right\} V. \end{aligned}$$

It is furthermore clear that  $\Delta(\beta_1, \beta_2, \dots, \beta_q)$  is a polynomial of degree  $n-1$  in each  $\beta_j$ . Since  $\Delta$  vanishes when any two  $\beta_i$  are equated,  $\Delta$  must have  $V$  as a factor. Hence the quotient

$$\Phi(\beta_1, \beta_2, \dots, \beta_q) = \frac{\Delta(\beta_1, \beta_2, \dots, \beta_q)}{V(\beta_1, \beta_2, \dots, \beta_q)}$$

is a polynomial of degree

$$(n-1) - (q-1) = n-q = p-1$$

in each  $\beta_j$ , and, as is evident from formulas (4.1) and (4.2), it is symmetric in the  $\beta_j$ .

Since  $P(z)$  is a polynomial of degree  $p$  and has no multiple zeros, it is true, according to Lagrange's interpolation formula that

$$\frac{\Phi(\beta_1, \beta_2, \dots, \beta_q)}{P(\beta_1)} = \sum_{i=1}^p \frac{\Phi(\alpha_{i1}, \beta_2, \beta_3, \dots, \beta_q)}{P'(\alpha_{i1})(\beta_1 - \alpha_{i1})}.$$

If, now, it be assumed that, for  $m$  any fixed positive integer less than  $q$ ,

$$\begin{aligned} & \frac{\Phi(\beta_1, \beta_2, \dots, \beta_q)}{P(\beta_1)P(\beta_2) \cdots P(\beta_m)} \\ &= \sum_{i_1=1}^p \sum_{i_2=1}^p \cdots \sum_{i_m=1}^p \frac{\Phi(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_m}, \beta_{m+1}, \beta_{m+2}, \dots, \beta_q)}{P'(\alpha_{i_1})P'(\alpha_{i_2}) \cdots P'(\alpha_{i_m})(\beta_1 - \alpha_{i_1})(\beta_2 - \alpha_{i_2}) \cdots (\beta_m - \alpha_{i_m})}, \end{aligned}$$

then again by Lagrange's formula

$$\begin{aligned} & \frac{\Phi(\beta_1, \beta_2, \dots, \beta_q)}{P(\beta_1)P(\beta_2) \cdots P(\beta_{m+1})} \\ &= \sum_{i_1=1}^p \sum_{i_2=1}^p \cdots \sum_{i_m=1}^p \frac{1}{P'(\alpha_{i_1})P'(\alpha_{i_2}) \cdots P'(\alpha_{i_m})(\beta_1 - \alpha_{i_1})(\beta_2 - \alpha_{i_2}) \cdots (\beta_m - \alpha_{i_m})} \\ & \quad \cdot \sum_{i_{m+1}=1}^p \frac{\Phi(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_{m+1}}, \beta_{m+2}, \dots, \beta_q)}{P'(\alpha_{i_{m+1}})(\beta_{m+1} - \alpha_{i_{m+1}})}. \end{aligned}$$

It follows then by mathematical induction that

$$\begin{aligned} & \frac{\Phi(\beta_1, \beta_2, \dots, \beta_q)}{P(\beta_1)P(\beta_2) \cdots P(\beta_q)} \\ (5.2) \quad &= \sum_{i_1=1}^p \sum_{i_2=1}^p \cdots \sum_{i_q=1}^p \frac{\Phi(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q})}{P'(\alpha_{i_1}) \cdots P'(\alpha_{i_q})(\beta_1 - \alpha_{i_1}) \cdots (\beta_q - \alpha_{i_q})}. \end{aligned}$$

Let us next compute the value of  $\Phi(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q})$  for a given set  $A$  of the  $\alpha_{i_i}$ .

First, let us consider the case that no two  $\alpha_{i_i}$  of set  $A$  are equal. Then from formula (4.2) it follows that

$$V(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q}) \neq 0;$$

and, since the  $\alpha_{i_i}$  are zeros of  $P(z)$ , it follows from (4.1) or (5.1) that

$$\Delta(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q}) = P'(\alpha_{i_1})P'(\alpha_{i_2}) \cdots P'(\alpha_{i_q})V(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q}) \neq 0$$

and hence that

$$\Phi(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q}) = P'(\alpha_{i_1})P'(\alpha_{i_2}) \dots P'(\alpha_{i_q}).$$

Secondly, let us consider the case that in the set  $A$ ,  $\alpha_{i_1} = \alpha_{i_2} = \dots = \alpha_{i_\mu}$ , ( $\mu \leq q$ ), but  $\alpha_{i_\mu}, \alpha_{i_{\mu+1}}, \dots, \alpha_{i_q}$  are distinct. From formulas (4.1) and (4.2) it then follows that the derivatives

$$(6.1) \quad \left[ \frac{\partial^{k_1}}{\partial \beta_1^{k_1}} \frac{\partial^{k_2}}{\partial \beta_2^{k_2}} \dots \frac{\partial^{k_\mu}}{\partial \beta_\mu^{k_\mu}} V \right]_A,$$

$$(6.2) \quad \left[ \frac{\partial^{k_1}}{\partial \beta_1^{k_1}} \frac{\partial^{k_2}}{\partial \beta_2^{k_2}} \dots \frac{\partial^{k_\mu}}{\partial \beta_\mu^{k_\mu}} \Delta \right]_A$$

vanish whenever two or more  $k_i$  are equal and, therefore, whenever

$$k_1 + k_2 + \dots + k_\mu \leq 0 + 1 + 2 + \dots + (\mu - 1) = \mu(\mu - 1)/2$$

unless  $(k_1, k_2, \dots, k_\mu)$  is the set  $K_1: (0, 1, \dots, \mu - 1)$  or a set obtainable by merely permuting the numbers of the set  $K_1$ . These  $\mu!$  sets will be referred to hereafter as the sets  $K$ .

In the neighborhood of the point  $A$ ,

$$\Phi(\beta_1, \beta_2, \dots, \beta_q)$$

$$\begin{aligned} & \sum \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_\mu^{k_\mu} \left\{ \left[ \frac{\partial^{k_1}}{\partial \beta_1^{k_1}} \frac{\partial^{k_2}}{\partial \beta_2^{k_2}} \dots \frac{\partial^{k_\mu}}{\partial \beta_\mu^{k_\mu}} \Delta \right]_A + \epsilon_{k_1 k_2 \dots k_q} \right\} \\ &= \frac{\sum \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_\mu^{k_\mu} \left\{ \left[ \frac{\partial^{k_1}}{\partial \beta_1^{k_1}} \frac{\partial^{k_2}}{\partial \beta_2^{k_2}} \dots \frac{\partial^{k_\mu}}{\partial \beta_\mu^{k_\mu}} V \right]_A + \epsilon'_{k_1 k_2 \dots k_q} \right\}}{\sum \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_\mu^{k_\mu} \left\{ \left[ \frac{\partial^{k_1}}{\partial \beta_1^{k_1}} \frac{\partial^{k_2}}{\partial \beta_2^{k_2}} \dots \frac{\partial^{k_\mu}}{\partial \beta_\mu^{k_\mu}} V \right]_A + \epsilon'_{k_1 k_2 \dots k_q} \right\}} \end{aligned}$$

where  $\zeta_i = \beta_i - \alpha_{i_i}$ ,

$$[\epsilon_{k_1 k_2 \dots k_q}]_A = [\epsilon'_{k_1 k_2 \dots k_q}]_A = 0,$$

and both sums are taken over all sets  $K$ . Furthermore, since changing from one set  $K$  to another set  $K$  merely multiplies both derivatives (6.1) and (6.2) by one or both by minus one, we may write

$$\Phi(\beta_1, \beta_2, \dots, \beta_q)$$

$$\begin{aligned} & \sum \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_\mu^{k_\mu} \left\{ \pm \left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} \Delta \right]_A + \epsilon_{k_1 k_2 \dots k_q} \right\} \\ &= \frac{\sum \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_\mu^{k_\mu} \left\{ \pm \left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} V \right]_A + \epsilon'_{k_1 k_2 \dots k_q} \right\}}{\sum \zeta_1^{k_1} \zeta_2^{k_2} \dots \zeta_\mu^{k_\mu} \left\{ \pm \left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} V \right]_A + \epsilon'_{k_1 k_2 \dots k_q} \right\}} \end{aligned}$$

If therefore for a given path of approach of the point  $(\beta_1, \beta_2, \dots, \beta_q)$  to  $A$   $\lim (\zeta_i/\zeta_1) = \eta_i$ , then

$$\begin{aligned}\Phi(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q}) &= \frac{\left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} \Delta \right]_A \sum (\pm) \eta_2^{k_2} \eta_3^{k_3} \dots \eta_\mu^{k_\mu}}{\left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} V \right]_A \sum (\pm) \eta_2^{k_2} \eta_3^{k_3} \dots \eta_\mu^{k_\mu}} \\ &= \frac{\left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} \Delta \right]_A}{\left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} V \right]_A}.\end{aligned}$$

Finally, according to formula (5.1),

$$\begin{aligned}\left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} \Delta \right]_A &= \sum_{i_1=0}^1 \sum_{i_2=0}^2 \dots \sum_{i_\mu=0}^\mu \sum_{i_{\mu+1}=0}^1 \dots \sum_{i_q=0}^1 C_{2,i_2} C_{3,i_3} \dots C_{\mu,i_\mu} \\ &\quad \cdot P^{(1-i_1)}(\alpha_{i_1}) P^{(2-i_2)}(\alpha_{i_2}) \dots P^{(\mu-i_\mu)}(\alpha_{i_\mu}) P^{(1-i_{\mu+1})}(\alpha_{i_{\mu+1}}) \\ &\quad \dots P^{(1-i_q)}(\alpha_{i_q}) \left[ \frac{\partial^{i_1}}{\partial \beta_1^{i_1}} \frac{\partial^{i_2}}{\partial \beta_2^{i_2}} \dots \frac{\partial^{i_q}}{\partial \beta_q^{i_q}} V \right]_A.\end{aligned}$$

Since the  $\alpha_{i_i}$  are zeros of  $P(x)$ ,

$$\begin{aligned}\left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} \Delta \right]_A &= \sum_{i_2=0}^1 \sum_{i_3=0}^2 \dots \sum_{i_\mu=0}^{\mu-1} C_{2,i_2} C_{3,i_3} \dots C_{\mu,i_\mu} \\ &\quad \cdot P'(\alpha_{i_1}) P^{(2-i_2)}(\alpha_{i_2}) \dots P^{(\mu-i_\mu)}(\alpha_{i_\mu}) P'(\alpha_{i_{\mu+1}}) \\ &\quad \dots P'(\alpha_{i_q}) \left[ \frac{\partial^{i_2}}{\partial \beta_2^{i_2}} \frac{\partial^{i_3}}{\partial \beta_3^{i_3}} \dots \frac{\partial^{i_\mu}}{\partial \beta_\mu^{i_\mu}} V \right]_A.\end{aligned}$$

By use of our above remarks on the vanishing of the derivative (6.1), this expression reduces further to

$$\begin{aligned}\left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} \Delta \right]_A &= C_{2,1} C_{3,2} \dots C_{\mu,\mu-1} P'(\alpha_{i_1}) P'(\alpha_{i_2}) \\ &\quad \dots P'(\alpha_{i_q}) \left[ \frac{\partial}{\partial \beta_2} \frac{\partial^2}{\partial \beta_3^2} \dots \frac{\partial^{\mu-1}}{\partial \beta_\mu^{\mu-1}} V \right]_A;\end{aligned}$$

and, consequently,

$$\Phi(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_q}) = \mu! P'(\alpha_{i_1}) P'(\alpha_{i_2}) \dots P'(\alpha_{i_q}).$$

Thirdly, let us consider the case that all of the  $\alpha_{i_i}$  in the set  $A$  are distinct except for  $\mu$  of the  $\alpha_{i_i}$  which are equal to one another, these  $\mu$  of the  $\alpha_{i_i}$  not being necessarily the first  $\mu$  of the  $\alpha_{i_i}$ . Then, due to the symmetry of the function  $\Phi(\beta_1, \beta_2, \dots, \beta_q)$  in the  $\beta_i$ , we see that the result obtained in the second case holds here also; namely,

Suppose, for example, that

$$\alpha_1 = \alpha_2 = \cdots = \alpha_t, \quad t \leq p,$$

but that  $\alpha_t, \alpha_{t+1}, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q$  are distinct. Then equation (3) may be written as

$$(7) \quad \sum_{j_1=t}^p \sum_{j_2=t}^p \cdots \sum_{j_q=t}^p \frac{E_{j_1 j_2 \cdots j_q}}{(\beta_1 - \alpha_{j_1})(\beta_2 - \alpha_{j_2}) \cdots (\beta_q - \alpha_{j_q})} = 0$$

where  $E_{j_1 j_2 \cdots j_q}$  is a constant which will now be determined. If  $\alpha_t$  occurs exactly  $\lambda$  times in the denominator of a given term of equation (7), for example, in the product

$$(8) \quad (\beta_1 - \alpha_t)(\beta_2 - \alpha_t) \cdots (\beta_\lambda - \alpha_t),$$

that term may be considered as the limit of the sum of all terms of equation (3) in the denominators of which occur the products

$$(\beta_1 - \alpha_{j_1})(\beta_2 - \alpha_{j_2}) \cdots (\beta_\lambda - \alpha_{j_\lambda})$$

where the  $\alpha_{j_i}$  are selected in all possible ways from the set

$$(\alpha_1, \alpha_2, \dots, \alpha_t).$$

Suppose  $\kappa_1 \alpha_1$ 's,  $\kappa_2 \alpha_2$ 's,  $\dots$ , and  $\kappa_t \alpha_t$ 's, where  $\kappa_j \geq 0$ , all  $j$ , and

$$(9) \quad \kappa_1 + \kappa_2 + \cdots + \kappa_t = \lambda$$

are selected. There are in (3)

$$\frac{\lambda!}{\kappa_1! \kappa_2! \cdots \kappa_t!}$$

terms which contain the chosen  $\alpha_t$  and, according to Theorem 1, each of these terms will have as a factor of the numerator coefficient  $D$  the product

$$\kappa_1! \kappa_2! \cdots \kappa_t!.$$

Hence, the factor  $\lambda!$  occurs in the numerator of the limit of the sum of such terms. The set of nonnegative integers  $(\kappa_1, \kappa_2, \dots, \kappa_t)$  may, in addition, be selected subject to the condition (9) in  $C_{t+\lambda-1, \lambda}$  ways. Hence, the factor corresponding to (8) in the numerator of the given term of (7) will be

$$\lambda! C_{t+\lambda-1, \lambda} = t(t+1)(t+2) \cdots (t+\lambda-1).$$

On the other hand, suppose that  $\beta_1 = \beta_2 = \cdots = \beta_u$  but that  $\beta_u, \beta_{u+1}, \dots, \beta_q$  are distinct. Then, since the number of terms of (3) in which  $\beta_1, \beta_2, \dots, \beta_u$  are associated with  $\delta_1 \alpha_1$ 's,  $\delta_2 \alpha_2$ 's,  $\dots$ ,  $\delta_p \alpha_p$ 's, where  $\delta_1 + \delta_2 + \cdots + \delta_p = u$ , is

$$\frac{u!}{\delta_1! \delta_2! \cdots \delta_p!},$$

that number of terms coalesce to form the single corresponding term of the limit of (3).

Thus the following corollary is evident:

**COROLLARY.** *Among the  $r+s$  distinct numbers*

$$A_1, A_2, \dots, A_r, \quad B_1, B_2, \dots, B_s$$

*let each  $A_i$  be a zero, of multiplicity at least  $p_i$ , of a polynomial  $f(z)$  of degree  $n$ , and each  $B_k$  a zero, of multiplicity at least  $q_k$ , of the derivative of  $f(z)$  where*

$$2 \leq p_1 + p_2 + \dots + p_r = p \leq n$$

*and*

$$q_1 + q_2 + \dots + q_s = q = n - p + 1.$$

*Then the  $A_i$  and  $B_k$  satisfy the relation*

$$\sum_{j=1}^s \prod_{k=1}^r \prod_{i=1}^{\mu_k} \frac{\mu_k! q_i! C_{p_k + \mu_k - 1, \mu_k}}{\nu_{jk}! (B_j - A_k)^{\nu_{jk}}} = 0$$

*where the sum is formed for all  $\nu_{jk}$ , ( $j=1, 2, \dots, s$ ;  $k=1, 2, \dots, r$ ), such that  $\nu_{jk}=0, 1, 2, \dots, q_j$  and  $\nu_{j1} + \nu_{j2} + \dots + \nu_{jr} = q_j$ , and where  $\mu_k = \nu_{1k} + \nu_{2k} + \dots + \nu_{sk}$ .*

**3. Proof of inequality (2).** Theorem 1 and its corollary will now be applied to the establishing of the following theorem:

**THEOREM 2.** *If a polynomial  $f(z)$  of degree  $n$ , ( $n \geq 2$ ), has  $p$ , ( $p \geq 2$ ), zeros in or on a circle  $K$  of radius  $R$ , then its derivative  $f'(z)$  has at least  $p-1$  zeros in or on the concentric circle  $K'$  of radius*

$$R' = R \csc \frac{\pi}{2(n-p+1)}.$$

For the proof of Theorem 2, it may be assumed without loss of generality that  $K$  is the unit circle  $|z|=1$ .

Let  $\alpha_1, \alpha_2, \dots, \alpha_p$  be the  $p$  given zeros of  $f(z)$ , and let  $\beta_1, \beta_2, \dots, \beta_{n-1}$  be all  $n-1$  zeros of  $f'(z)$ , the subscripts on the  $\beta_j$  being chosen so that

$$|\beta_1| \geq |\beta_2| \geq \dots \geq |\beta_{n-1}|.$$

If  $|\beta_q| \leq 1$ , where  $q = n - p + 1$ , then also  $|\beta_j| \leq 1$ , ( $j = q+1, q+2, \dots, n-1$ ); that is to say, at least  $(n-q) = (p-1)$  of the  $\beta_j$  will lie in the unit circle and therefore in the circle  $K'$ .

If  $|\beta_q| > 1$ , then likewise  $|\beta_j| > 1$ , ( $j = 1, 2, \dots, q-1$ ). As  $|\alpha_k| \leq 1$  for all  $k$ , no  $\beta_j$ , ( $j = 1, 2, \dots, q$ ), will be an  $\alpha_k$ ; hence either Theorem 1 or its corol-

lary may be used. Let  $\phi_i$  be the angle subtended by the circle  $K$  in the point  $\beta_i$ , and let  $\alpha'_i$  denote the  $\alpha_k$  corresponding to a given  $\beta_i$  such that

$$0 \leq \arg \frac{\beta_i - \alpha'_i}{\beta_i - \alpha_k} \leq \phi_i \leq \phi_q < \pi$$

for all  $j=1, 2, \dots, q$  and all  $k=1, 2, \dots, p$ . It follows that

$$0 \leq \arg \frac{\prod_{j=1}^q (\beta_j - \alpha'_j)}{\prod_{j=1}^q (\beta_j - \alpha_{k_j})} \leq (q - \delta)\phi_q$$

where  $\delta$ , ( $0 \leq \delta \leq q$ ), denotes the number of factors common to the two products

$$\prod_{j=1}^q (\beta_j - \alpha'_j), \quad \prod_{j=1}^q (\beta_j - \alpha_{k_j}).$$

If, therefore,  $\phi_q < \pi/q$ , each term in the sum obtained on multiplying the left-hand side of (3) by

$$\prod_{j=1}^q (\beta_j - \alpha'_j)$$

could be represented by a vector drawn from the origin to a point lying in the angular opening

$$0 \leq \arg z < \pi;$$

hence the left-hand side of (3) would not vanish. As this result would contradict Theorem 1, it follows that  $\phi_q \geq \pi/q$ ; that is to say, the  $p-1$  zeros of  $f'(z)$   $\beta_q, \beta_{q+1}, \dots, \beta_{n-1}$  lie in or on a circle  $K'$  concentric with  $K$  and of radius

$$R' = \csc \frac{\pi}{2q} = \csc \frac{\pi}{2(n-p+1)}.$$

The above method of proof may also be used, with little change, in the case that  $K$  is a convex region not necessarily a circle. The corresponding result may be stated as follows:

**THEOREM 2'.** *If a polynomial  $f(z)$  of degree  $n$ , ( $n \geq 2$ ), has  $p$ , ( $p \geq 2$ ), zeros in a convex region  $K$ , its derivative has at least  $p-1$  zeros in the star-shaped region  $K'$  consisting of all points of the plane from which  $K$  subtends an angle of not less than  $\pi/(n-p+1)$  radians.*

Theorem 2 or Theorem 2' does not furnish, however, the least number

$\rho(n, p)$  as defined in §1. This is clear from the fact that, in general, the quantity  $\delta$  used in the proof takes on values in addition to 0 and that, therefore  $\phi_q$  must be actually greater than  $\pi/q$  in order for the left-hand side of (3) to vanish.

The same is clear from the facts that, although for  $p = n$

$$\csc \frac{\pi}{2(n-p+1)} = 1 = \rho(n, n),$$

nevertheless for  $p = 2$  and  $n \geq 3$

$$\csc \frac{\pi}{2(n-p+1)} > \csc \pi/n \geq \rho(n, 2),$$

and for\*  $p = n - 1$  and  $n \geq 2$

$$\csc \frac{\pi}{2(n-p+1)} = 2^{1/2} > (1 + 1/n)^{1/2} \geq \rho(n, n-1).$$

#### 4. Relation to a theorem of Fekete. The inequality

$$\rho(n, 2) \leq \csc \pi/n$$

was proved by Szegő as a consequence of the following theorem of Grace and Heawood:† If  $a$  and  $b$  are two distinct zeros of a polynomial  $f(z)$  of degree  $n$ , at least one zero of the derivative of  $f(z)$  lies in or on the circle

$$(10) \quad \left| z - \frac{a+b}{2} \right| = \left| \frac{a-b}{2} \right| \cot \pi/n.$$

Szegő showed that if the  $a$  and  $b$  are allowed to vary independently in and on the unit circle, the envelope of circle (10) is the circle  $|z| = \csc \pi/n$ .

A similar relation will now be proved to hold between Theorem 2 and the following theorem:

**THEOREM 3.** *If  $a$  and  $b$  are respectively  $k$ -fold and  $l$ -fold zeros of a polynomial  $f(z)$  of degree  $n$ , then at least one zero (different from  $a$  and  $b$ ) of the derivative lies in or on the circle*

$$(11) \quad \left| z - \frac{a+b}{2} \right| = \left| \frac{a-b}{2} \right| \cot \frac{\pi}{2(n+1-k-l)}.$$

\* For  $p = n - 1$  and  $n \geq 5$ ,  $\csc \pi/[2(n-p+1)] = 2^{1/2} > 1 + 2/n$ , where  $1 + 2/n$  is a limit obtainable from a theorem due to Walsh. See J. L. Walsh, these Transactions, vol. 24 (1922), p. 37, and also Biernacki, Bulletin de l'Académie Polonaise, 1927, p. 121.

† J. H. Grace, Proceedings of the Cambridge Philosophical Society, vol. 11 (1901), pp. 352-357; P. J. Heawood, Quarterly Journal of Mathematics, vol. 38 (1907), pp. 84-107.

This theorem, a generalization of one due to Fekete,\* is an immediate result of the lemma:†

If  $P(z)$  is a polynomial of degree  $\nu \geq 1$ , if  $\phi(z)$  is a function real, continuous, nonnegative, and not identically vanishing on the interval  $(-1, 1)$  of the real axis, and if

$$\int_{-1}^1 \phi(z)P(z)dz = 0,$$

then  $P(z)$  vanishes in at least one point in which the segment  $(-1, 1)$  subtends an angle of not less than  $\pi/\nu$ .

In the proof of Theorem 3, it may, without loss of generality, be assumed that  $a = -1$  and  $b = 1$ . If  $\phi(z) = (1+z)^{k-1}(1-z)^{l-1}$  and  $P(z) = f'(z)/\phi(z)$ , the latter being a polynomial of degree  $\nu = n+1-k-l$ , the requirements of the lemma just quoted will be satisfied and Theorem 3 will follow at once.

It will now be shown that the envelope of the circles (11) when  $a$  and  $b$  vary independently in or on the unit circle is the circle of Theorem 2 with  $p = k+l$ . It obviously suffices to find the envelope of the circles (11) when  $a$  and  $b$  vary on the unit circle. Every point of circle (11) may then have its coordinates written in the form

$$z = \frac{a+b}{2} + \theta \left( \frac{a-b}{2} \right) \cot \frac{\pi}{2(n+1-k-l)}$$

with  $|\theta| \leq 1$ , and  $|a| = |b|$ . An angle  $\psi$  may be found so that either  $a = be^{i\psi}$  or  $b = ae^{i\psi}$  where  $0 \leq \psi \leq \pi$ . In either case

$$\begin{aligned} |z| &\leq \cos \frac{\psi}{2} + \sin \frac{\psi}{2} \cot \frac{\pi}{2(n+1-k-l)} \\ &= \frac{\sin \left[ \frac{\psi}{2} + \frac{\pi}{2(n+1-k-l)} \right]}{\sin \frac{\pi}{2(n+1-k-l)}} \leq \csc \frac{\pi}{2(n+1-k-l)}. \end{aligned}$$

**5.  $p$ -valent polynomials.** An immediate corollary of Theorem 2 is the theorem:‡

\* M. Fekete, *Acta Litterarum ac Scientiarum*, Szeged, vol. 1 (1923), pp. 98-100.

† M. Fekete, *Mathematische Zeitschrift*, vol. 22 (1925), p. 2, and *Jahresbericht der deutschen Mathematiker-Vereinigung*, vol. 34 (1926), p. 211. See also M. Marden, *Bulletin of the American Mathematical Society*, vol. 38 (1932), p. 440; vol. 39 (1933), pp. 750-754.

‡ Alexander and Kakeya gave this theorem in the special case  $p = 1$ . See the above references.

If the derivative of a polynomial  $f(z)$  of degree  $n \geq 2$  has exactly  $p-1$  zeros ( $2 \leq p < n$ ) in the unit circle, then  $f(z)$  has at most  $p$  zeros in or on the circle

$$|z| = \sin \frac{\pi}{2(n-p)}.$$

For, if  $f(z)$  had  $p+1$  zeros in this circle,  $f'(z)$  would have at least  $p$  zeros in or on the circle

$$|z| = \sin \frac{\pi}{2(n-p)} \csc \frac{\pi}{2(n-p)}$$

in contradiction to the hypothesis.

This corollary is essentially identical with the following theorem about  $p$ -valent polynomials:

**THEOREM 4.** *If the derivative of a polynomial  $P(z)$  of degree  $n$ , ( $n \geq 2$ ), has exactly  $p-1$  zeros ( $2 \leq p < n$ ) in or on the unit circle, then  $P(z)$  is at most  $p$ -valent in or on the circle*

$$|z| = \sin \frac{\pi}{2(n-p)}.$$

By a function's being  $p$ -valent in a given region  $R$  it is meant that the function takes on at least one value  $p$  times in  $R$  and no value more than  $p$  times in  $R$ . It suffices then merely to set  $f(z) = P(z) - \gamma$ , where  $\gamma$  is an arbitrary constant, in order to deduce Theorem 4 from the above corollary.

Finally, the same method of reasoning when used together with Theorem 2' leads to the following more general conclusion giving a sufficient condition for a polynomial to be at most  $p$ -valent in a convex region  $K$ , not necessarily a circle.

**THEOREM 4'.** *Let  $K$  be a convex region and  $S$  the star-shaped region comprised of all points from which  $K$  subtends an angle of at least  $\pi/(n-p)$  radians ( $2 \leq p < n$ ). Then, if the derivative of any polynomial  $P(z)$  of the  $n$ th degree has exactly  $p-1$  zeros in  $S$ , the polynomial  $P(z)$  is at most  $p$ -valent in  $K$ .*

UNIVERSITY OF WISCONSIN AT MILWAUKEE,  
MILWAUKEE, WIS.

## ABSTRACT SYMMETRIC BOUNDARY CONDITIONS\*

BY

J. W. CALKIN

The topic discussed here belongs to the theory of linear transformations in Hilbert space and presupposes on the part of the reader a fairly thorough knowledge of certain portions of that subject. The following material is at most only slightly more than the minimum prerequisite: M. H. Stone, *Linear Transformations in Hilbert Space and their Applications to Analysis*, American Mathematical Society Colloquium Publications, vol. 15, New York, 1932, chaps. 1-5, and chap. 9, §§1, 2, or the various writings of J. von Neumann dealing with the same or related aspects of the theory; J. von Neumann, *Über adjungierte Operatoren*, Annals of Mathematics, (2), vol. 33 (1932), pp. 294-310; F. J. Murray, *Linear transformations between Hilbert spaces and the application of this theory to linear partial differential equations*, these Transactions, vol. 37 (1935), pp. 301-338, §§1-5 only.

I wish here to acknowledge my indebtedness to M. H. Stone who not only suggested the thesis indicated above, but has also made an extremely valuable contribution to the present work. Thanks are due also to J. von Neumann, with whom the author has had several fruitful conversations concerning the theory here developed. More precise acknowledgments are made in the course of the paper.

### INTRODUCTION

1. **The nature and applications of the subject.** The basic concept of the present paper is embodied in a definition (Definition 1.1) which leads to a formula associated with a certain type of transformation  $T$  in Hilbert space, analogous to the so-called "fundamental formula" associated with a differential operator coincident with its formal adjoint.† We are thus able to introduce an abstract definition of linear boundary conditions associated with the equation  $Tf - \lambda f = g$ , and to study the properties of such boundary conditions.

\* Presented (in part) to the Society, February 26, 1938; received by the editors July 21, 1938.

Several of the theorems of Chapters III and IV are abstract formulations of results which appeared in the writer's doctoral thesis, *Applications of the Theory of Hilbert Space to Partial Differential Equations*, Harvard, 1937 (see abstracts 43-3-114, 43-3-209, Bulletin of the American Mathematical Society).

† We use the term "fundamental formula" in the sense of Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, New Haven, 1923, pp. 58-69; for an ordinary differential operator, the fundamental formula is the familiar Lagrange identity.

In this formulation, the boundary condition is regarded as defining a subset of the domain of  $T$  and thus a contraction of the transformation  $T$  (that is, a transformation  $S$  such that  $S \subseteq T$ ). The main problem which we consider is the determination of those boundary conditions which define self-adjoint and maximal symmetric contractions of  $T$ .

The entities introduced in the fundamental definition are realizable in terms of a wide variety of differential operators, both ordinary and partial. In terms of such realizations, the abstract boundary conditions which we consider include the familiar self-adjoint boundary conditions of the classical theory of differential equations. Applied to a differential operator our results serve to characterize a wide class of differential systems for which a unique spectral form exists.

Apart from the introduction, the paper is divided into four chapters. In Chapter I the fundamental definitions are introduced and a few simple but basic theorems established. Here also examples from the field of differential operators are given. The chapter concludes with a precise statement of the important problems to be considered. In Chapter II, manifolds possessing a kind of symmetry, including as special cases manifolds which appear as the graphs of symmetric transformations, are studied in detail. In Chapter III, the situation postulated in Definition 1.1 which gives rise to the "fundamental formula" for an operator  $T$  is thoroughly analyzed. In Chapter IV, the results of Chapters II and III are applied to the solution of the problems stated at the end of Chapter I.

A fifth chapter dealing with the applications of the theory to certain types of differential operators was originally planned but is not included; applications will be considered in subsequent papers.

For the convenience of the reader, a detailed table of contents appears at the end of the introduction.

**2. Notation, terminology, and conventions.** Except for minor modification and additions, we use the notation and terminology of M. H. Stone.\*

We take occasion here to point out the following notations which we employ systematically and which are not entirely standardized:  $\mathfrak{D}$  for a unitary space with dimension number zero and especially for the subspace with dimension number zero of any space under consideration;  $\mathfrak{D}(T)$  and  $\mathfrak{R}(T)$  for the domain and range, respectively, of a transformation  $T$ ;  $T\mathfrak{N}$  for the set in the range of  $T$  into which  $T$  takes the set  $\mathfrak{N}$  in its domain. We reserve the letter  $E$  for the designation of projections, and denote a projection with range

---

\* *Linear Transformations in Hilbert Space and Their Applications to Analysis*. All citations of Stone in the sequel refer to this book.

$\mathfrak{N}$  by  $E_{\mathfrak{N}}$ . We find it convenient also to adopt the topological notation  $\bar{\mathfrak{N}}$  for the closure of the set  $\mathfrak{N}$ . Since we use this notation only where  $\mathfrak{N}$  is a linear manifold,  $\bar{\mathfrak{N}}$  is precisely the closed linear manifold determined by  $\mathfrak{N}$ .

We employ extensively the concept of the graph of a transformation. The graph of a transformation  $T$  with domain in a Hilbert space  $\mathfrak{H}_1$  and range in a Hilbert space  $\mathfrak{H}_2$  is the manifold of vectors  $\{f, Tf\}$  in the space  $\mathfrak{H}_1 \oplus \mathfrak{H}_2$ ; in particular,  $\mathfrak{H}_1$  and  $\mathfrak{H}_2$  may be identical.<sup>†</sup> We also admit the possibility that either  $\mathfrak{H}_1$  or  $\mathfrak{H}_2$  is a unitary space. Except where otherwise indicated, we denote the graph of a transformation  $T$  by the symbol  $\mathfrak{B}(T)$ . We recall that  $\mathfrak{B}(T)$  is a closed linear manifold if and only if  $T$  is a closed linear transformation; and that, if  $T^*$  exists,  $(\mathfrak{H}_1 \oplus \mathfrak{H}_2) \ominus \mathfrak{B}(T)$  is the linear manifold of vectors  $\{T^*f, -f\}$ .

At various points we must discuss questions involving a space which may be either a Hilbert space or a unitary space—that is, a separable complex Euclidean space—and we use the terminology ordinarily associated with the theory of transformations in Hilbert space to cover both cases. This necessitates our taking the definitions of various types of transformations in Hilbert space as definitions of transformations in unitary space also. In many cases the distinctions which these definitions set up for transformations in Hilbert space are vacuous for transformations in a space with finite dimension number. For example, in a unitary space, every linear transformation is bounded, every linear symmetric transformation is self-adjoint, every maximal isometric transformation is unitary. These facts, however, are all well known and in many cases self-evident. We do not, therefore, make explicit comment at every point in the sequel where specialization to the case of unitary space makes modification of the exposition possible.

Although we have occasion to discuss mathematical relations involving several inner products, not all formed in the same space, we use the same notation, namely  $(,)$ , for all inner products under consideration and state in which space each is formed only when the context fails to make it clear.

In discussing the orthogonal sum,  $\mathfrak{H} = \mathfrak{H}_1 \oplus \mathfrak{H}_2 \oplus \cdots \oplus \mathfrak{H}_n$ , of a finite collection of spaces, we find it convenient to use the notation  $\mathfrak{H}_k$  to mean, as well as the space  $\mathfrak{H}_k$  itself, that manifold of vectors in  $\mathfrak{H}$  whose components in  $\mathfrak{H}_j$ , ( $j \neq k$ ), are all zero. More generally, if  $\mathfrak{M}$  denotes a manifold in  $\mathfrak{H}_{k_1} \oplus \mathfrak{H}_{k_2} \oplus \cdots \oplus \mathfrak{H}_{k_m}$ , ( $m < n$ ), we shall on occasion use the symbol  $\mathfrak{M}$  also to mean that manifold in  $\mathfrak{H}$  whose projection on  $\mathfrak{H}_{k_1} \oplus \mathfrak{H}_{k_2} \oplus \cdots \oplus \mathfrak{H}_{k_m}$  is  $\mathfrak{M}$ .

<sup>†</sup> The notion of the graph for the case  $\mathfrak{H}_1 = \mathfrak{H}_2$  is due to J. von Neumann, *Annals of Mathematics*, (2), vol. 33 (1932), pp. 294–310; especially p. 299. The more general definition was introduced by F. J. Murray, these *Transactions*, vol. 37 (1935), pp. 301–338; especially pp. 302–303. All future citations of Murray refer to this paper.

and whose projection on  $\mathfrak{S}_j$ , ( $j \neq k_1, k_2, \dots, k_m$ ), is zero. However, when there is danger of ambiguity, we use a different convention. Thus, if the space under consideration is  $\mathfrak{S} \oplus \mathfrak{S}$  and  $\mathfrak{M}$  is a manifold in  $\mathfrak{S}$ , we shall use the notation  $\mathfrak{M} + \mathfrak{D}$  to denote the manifold of vectors  $\{f, 0\}$  in  $\mathfrak{S} \oplus \mathfrak{S}$  such that  $f$  is in  $\mathfrak{M}$ .

## CONTENTS

### Chapter I. Fundamental concepts

1. Reduction operators.....	372
2. Linear boundary conditions.....	377
3. The fundamental problem.....	377

### Chapter II. $W$ -symmetric manifolds

1. Isometric transformations.....	378
2. $W$ -symmetric manifolds and isometric transformations.....	379
3. Illustrations and applications.....	383
4. Real $W$ -symmetric manifolds.....	389

### Chapter III. Reduction operators

1. An alternative definition.....	391
2. Characterization of all reduction operators.....	392
3. The graph of a reduction operator.....	396
4. Two types of reduction operator.....	400
5. Self-adjoint extension by means of reduction operators.....	404

### Chapter IV. Boundary conditions

1. Introduction.....	406
2. Boundary conditions involving a bounded reduction operator.....	406
3. Unbounded reduction operators; preliminary questions.....	409
4. Boundary conditions involving an arbitrary reduction operator.....	416
5. Pathology of unbounded reduction operators.....	424
6. A special class of boundary conditions.....	434
7. Real boundary conditions.....	437
8. Formulation of boundary conditions.....	439
9. Homogeneous boundary value problems.....	440

## CHAPTER I. FUNDAMENTAL CONCEPTS

1. **Reduction operators.** The basic notion of this paper, indicated roughly in the introduction, we now state precisely.

**DEFINITION 1.1.** Let  $H$  be a closed linear transformation in a Hilbert space  $\mathfrak{S}$ , and let  $H^*$  exist. A transformation  $A$  with domain in the graph of  $H^*$  and

with range in a unitary or Hilbert space  $\mathfrak{M}$  is said to be a reduction operator for  $H^*$  if the following conditions are satisfied:

- (1)  $A$  is closed, linear, and has domain dense in  $\mathfrak{B}(H^*)$ ;
- (2) there exists a unitary transformation  $W$  in  $\mathfrak{M}$  such that

$$(\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}) \ominus \mathfrak{B}(A)$$

is the set of all vectors  $\{H^*f, -f, WA\{f, H^*f\}\}$ .

The space  $\mathfrak{M}$  is called the range-space of  $A$ .†

It may be observed that Definition 1.1 can be modified as follows: The condition that  $A$  be closed linear can be omitted from (1) and the condition (2) can be stated in a form which does not require  $\mathfrak{B}(A)$  to be a closed linear manifold. It can then be proved that  $A$  is necessarily closed and linear.

From Definition 1.1, we have at once the formula

$$(1.1) \quad (f, H^*g) - (H^*f, g) + (A\{f, H^*f\}, WA\{g, H^*g\}) = 0,$$

for all  $f$  and  $g$  in  $\mathfrak{B}(H^*)$  such that  $A\{f, H^*f\}$  and  $A\{g, H^*g\}$  are defined. In order to simplify the notation, we shall hereafter often write  $Af$  for  $A\{f, H^*f\}$ . Thus the abstract "Lagrange identity" (1.1) may be written

$$(1.2) \quad (f, H^*g) - (H^*f, g) = -(Af, WAf).$$

**THEOREM 1.1.** *The transformation  $H$  is symmetric. The domain of  $A$  contains the graph of  $H$ , and  $A\{f, H^*f\} = 0$  if and only if  $f$  is in the domain of  $H$ . Thus  $\mathfrak{D}(A) = \mathfrak{B}(H)$  if and only if  $H$  is self-adjoint.*

Since  $H^*$  exists and  $H$  is linear,  $H$  has domain dense in  $\mathfrak{S}$ .‡ Moreover, since  $H$  is closed,  $H^*$  has domain dense in  $\mathfrak{S}$ , and  $H^{**}$  exists and is identically  $H$ .§

If  $A\{f, H^*f\} = 0$ , then  $\{H^*f, -f, 0\}$  is in  $(\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}) \ominus \mathfrak{B}(A)$ , by Definition 1.1, (2). Thus  $(g, H^*f) - (H^*g, f) = 0$  for all  $\{g, H^*g\}$  in  $\mathfrak{D}(A)$ . Since  $\mathfrak{D}(A)$  is dense in  $\mathfrak{B}(H^*)$ , it follows from the identity  $H \equiv H^{**}$  that  $f$  is in  $\mathfrak{D}(H)$  and that  $H^*f = Hf$ . On the other hand, if  $f$  is in  $\mathfrak{D}(H)$  it follows, again from the identity  $H \equiv H^{**}$ , that  $\{Hf, -f, 0\}$  is in  $(\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}) \ominus \mathfrak{B}(A)$ . Therefore  $\{f, Hf\}$  is in  $\mathfrak{D}(A)$  and  $A\{f, Hf\} = 0$ . Thus  $\mathfrak{D}(A) \supseteq \mathfrak{B}(H)$  and  $A\mathfrak{B}(H) = \mathfrak{D}$ . Moreover, since  $\mathfrak{B}(H^*) \supseteq \mathfrak{D}(A)$ , we have  $\mathfrak{B}(H^*) \supseteq \mathfrak{B}(H)$ . Hence  $H^* \supseteq H$ , and, since  $\mathfrak{D}(H)$  determines  $\mathfrak{S}$ ,  $H$  is symmetric.

Finally, since  $\mathfrak{D}(A)$  is dense in  $\mathfrak{B}(H^*)$  and  $H$  is closed, the equations

† Compare our previous definition, Proceedings of the National Academy of Sciences, vol. 24 (1938), pp. 38-42, Definition 1. That the requirement  $W^2 + I = 0$  is unnecessary was pointed out to us by J. von Neumann (cf. Theorem 1.2 below).

‡ Stone, Theorem 2.6.

§ Von Neumann, loc. cit., Theorem 2.

$\mathfrak{D}(A) = \mathfrak{B}(H)$  and  $\mathfrak{B}(H^*) = \mathfrak{B}(H)$  are clearly coextensive. Hence the identities  $A \equiv 0$  and  $H^* \equiv H$  are also.

**THEOREM 1.2.** *The range of  $A$  is dense in  $\mathfrak{M}$ , and  $W$  satisfies the identity  $W^2 + I \equiv 0$ .*

Since  $A$  is linear,  $\mathfrak{R}(A)$  is a linear manifold and is therefore dense in  $\mathfrak{M}$  if and only if  $\mathfrak{M} \ominus \overline{\mathfrak{R}(A)} = \mathfrak{D}$ . To establish the latter identity we have only to observe that if  $k$  belongs to  $\mathfrak{M} \ominus \overline{\mathfrak{R}(A)}$ , then

$$(f, 0) - (H^*f, 0) + (Af, k) = 0$$

for all  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$ , whence it follows that  $k = WA\{0, 0\} = 0$ .

To prove the second assertion of the theorem we set  $g = f$  in the identity (1.2) to obtain

$$(f, H^*f) - (H^*f, f) = -(Af, WAf).$$

Taking the complex conjugate of both members of this equation, we have

$$(H^*f, f) - (f, H^*f) = -(WAF, Af).$$

Thus, for  $h$  in  $\mathfrak{R}(A)$ , we obtain by addition of the two preceding equations the relation  $(Wh, h) + (h, Wh) = 0$ . Moreover, since  $W$  is unitary,  $(h, Wh) = (W^{-1}h, h)$  and thus  $(Wh, h) + (W^{-1}h, h) = 0$  for all  $h$  in  $\mathfrak{R}(A)$ . Hence, since  $W$  and  $W^{-1}$  are bounded and  $\mathfrak{R}(A) = \mathfrak{M}$ , we have  $(Wh + W^{-1}h, h) = 0$  for all  $h$  in  $\mathfrak{M}$ . But  $W + W^{-1}$  is self-adjoint; therefore the result just obtained implies that its bound is zero. Consequently  $W + W^{-1} \equiv 0^\dagger$  or  $W^2 + I \equiv 0$ , as we wished to prove.

**THEOREM 1.3.** *Let  $A$  be a reduction operator for  $H^*$ , and let  $T$  be an arbitrary bounded self-adjoint transformation in  $\mathfrak{S}$ . Let  $C$  be the transformation which has as its domain the set of elements  $\{f, (H^* + T)f\}$  of  $\mathfrak{B}(H^* + T)$  such that  $\{f, H^*f\}$  is in  $\mathfrak{D}(A)$ , and which takes  $\{f, (H^* + T)f\}$  into  $A\{f, H^*f\}$ . Then  $C$  is a reduction operator for  $H^* + T$ .*

To prove that  $C$  is a reduction operator for  $H^* + T$ , we seek all elements  $\{g^*, g, h\}$  of  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$  such that

$$(f, g^*) - ((H^* + T)f, g) + (Af, h) = 0$$

for all  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$ . Since  $\mathfrak{D}(A) \supseteq \mathfrak{B}(H)$  and  $A\mathfrak{B}(H) = \mathfrak{D}$ , the identity  $(H + T)^* \equiv H^* + T$ , which holds by virtue of the fact that  $T$  is bounded, implies that  $g$  is in  $\mathfrak{D}(H^*) = \mathfrak{D}(H^* + T)$  and that  $g^* = (H^* + T)g$ . Hence, since  $T$  is self-adjoint,  $\{g^*, g, h\}$  satisfies the above equation for all  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$  if and only if

<sup>†</sup> Stone, Theorem 2.22.

$$(f, H^*g) - (H^*f, g) + (Af, h) = 0$$

for all such  $\{f, H^*f\}$ . But the latter condition is satisfied if and only if  $h = WAg$ . Thus, in view of the remark following Definition 1.1 we conclude that  $C$  is a reduction operator for  $H^* + T$ .

To reveal more clearly the significance of Definition 1.1, we give now some concrete examples of reduction operators.

EXAMPLE 1.  $\mathfrak{S}$  is the space  $\mathfrak{L}_2(a, b)$  where  $a \leq x \leq b$  is a finite interval of the real axis.  $H^*$  is the transformation which takes  $f$  into  $if'$  and has for its domain  $\mathfrak{D}^*$  the set of all elements  $f$  of  $\mathfrak{L}_2(a, b)$  which are absolutely continuous on  $a \leq x \leq b$  and such that  $\int_a^b |f'|^2 dx < \infty$ . The operator  $H$  is the contraction of  $H^*$ , whose domain  $\mathfrak{D}$  is the set of all elements in  $\mathfrak{D}^*$  which vanish at  $a$  and  $b$ . The space  $\mathfrak{M}$  is a two-dimensional unitary space, and  $A$  is the transformation which takes  $\{f, H^*f\}$  into the point of  $\mathfrak{M}$  with coordinates  $(f(b), f(a))$ ;  $W$  is the transformation which takes the point  $(c, d)$  in  $\mathfrak{M}$  into  $(ic, -id)$ . Here the formula (1.1) is the formula of Lagrange,

$$\int_a^b f \overline{ig'} dx - \int_a^b if' \overline{g} dx + f(b) \overline{ig(b)} - f(a) \overline{ig(a)} = 0.$$

EXAMPLE 2.  $\mathfrak{S}$  is the same as in Example 1; the domain  $\mathfrak{D}^*$  of  $H^*$  is the set of all elements  $f$  of  $\mathfrak{L}_2(a, b)$  such that  $f$  and  $f'$  are absolutely continuous on  $a \leq x \leq b$  and  $\int_a^b |f''|^2 dx$  is finite;  $H^*f = f''$ . The domain  $\mathfrak{D}$  of  $H$  consists of those and only those elements  $g$  of  $\mathfrak{D}^*$  such that  $g(a) = g(b) = g'(a) = g'(b) = 0$ ;  $Hg = g''$ . The space  $\mathfrak{M}$  is a four-dimensional unitary space;

$$A\{f, H^*f\} = \{f(b), f(a), f'(a), -f'(b)\}.$$

The transformation  $W$  takes  $(h, k, l, m)$  into  $(m, l, -k, -h)$  and the formula (1.1) is

$$\int_a^b f \overline{g''} dx - \int_a^b \overline{g} f'' dx - f(b) \overline{g'(b)} + f(a) \overline{g'(a)} - f'(a) \overline{g(a)} + f'(b) \overline{g(b)} = 0.$$

EXAMPLE 3.  $\mathfrak{S}$  is the space  $\mathfrak{L}_2(E)$  where  $E$  is the set in the  $(x, y)$ -plane bounded by the lines  $x = a, x = b, y = c, y = d, a < b, c < d$ . The domain  $\mathfrak{D}^*$  of  $H^*$  is the set of all elements  $f(x, y)$  of  $\mathfrak{L}_2(E)$  which are absolutely continuous on  $a \leq x \leq b$  for almost all  $y$  on  $c < y < d$  and for which  $\int_E |f_x|^2 dE$  is finite;  $H^*f = if_x$ . The domain  $\mathfrak{D}$  of  $H$  is the set of all elements  $g$  of  $\mathfrak{D}^*$  such that  $\lim_{x \rightarrow a} g(x, y) = \lim_{x \rightarrow b} g(x, y) = 0$  for almost all  $y$  on  $c < y < d$ ;  $Hg = ig_x$ . The space  $\mathfrak{M}$  is the space  $\mathfrak{L}_2(c, d) \oplus \mathfrak{L}_2(c, d)$ ;

$$Af = \{f(b, y), f(a, y)\}; \quad W\{h(y), k(y)\} = -i\{h(y), -k(y)\}.$$

$\mathfrak{D}(A) = \mathfrak{B}(H)$  and  $\mathfrak{B}(H^*) = \mathfrak{B}(H)$  are clearly coextensive. Hence the identities  $A=0$  and  $H^*=H$  are also.

**THEOREM 1.2.** *The range of  $A$  is dense in  $\mathfrak{M}$ , and  $W$  satisfies the identity  $W^2 + I = 0$ .*

Since  $A$  is linear,  $\mathfrak{R}(A)$  is a linear manifold and is therefore dense in  $\mathfrak{M}$  if and only if  $\mathfrak{M} \ominus \overline{\mathfrak{R}(A)} = \mathfrak{D}$ . To establish the latter identity we have only to observe that if  $k$  belongs to  $\mathfrak{M} \ominus \overline{\mathfrak{R}(A)}$ , then

$$(f, 0) - (H^*f, 0) + (Af, k) = 0$$

for all  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$ , whence it follows that  $k = WA\{0, 0\} = 0$ .

To prove the second assertion of the theorem we set  $g = f$  in the identity (1.2) to obtain

$$(f, H^*f) - (H^*f, f) = - (Af, WAf).$$

Taking the complex conjugate of both members of this equation, we have

$$(H^*f, f) - (f, H^*f) = - (WAf, Af).$$

Thus, for  $h$  in  $\mathfrak{R}(A)$ , we obtain by addition of the two preceding equations the relation  $(Wh, h) + (h, Wh) = 0$ . Moreover, since  $W$  is unitary,  $(h, Wh) = (W^{-1}h, h)$  and thus  $(Wh, h) + (W^{-1}h, h) = 0$  for all  $h$  in  $\mathfrak{R}(A)$ . Hence, since  $W$  and  $W^{-1}$  are bounded and  $\mathfrak{R}(A) = \mathfrak{M}$ , we have  $(Wh + W^{-1}h, h) = 0$  for all  $h$  in  $\mathfrak{M}$ . But  $W + W^{-1}$  is self-adjoint; therefore the result just obtained implies that its bound is zero. Consequently  $W + W^{-1} = 0^\dagger$  or  $W^2 + I = 0$ , as we wished to prove.

**THEOREM 1.3.** *Let  $A$  be a reduction operator for  $H^*$ , and let  $T$  be an arbitrary bounded self-adjoint transformation in  $\mathfrak{S}$ . Let  $C$  be the transformation which has as its domain the set of elements  $\{f, (H^* + T)f\}$  of  $\mathfrak{B}(H^* + T)$  such that  $\{f, H^*f\}$  is in  $\mathfrak{D}(A)$ , and which takes  $\{f, (H^* + T)f\}$  into  $A\{f, H^*f\}$ . Then  $C$  is a reduction operator for  $H^* + T$ .*

To prove that  $C$  is a reduction operator for  $H^* + T$ , we seek all elements  $\{g^*, g, h\}$  of  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$  such that

$$(f, g^*) - ((H^* + T)f, g) + (Af, h) = 0$$

for all  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$ . Since  $\mathfrak{D}(A) \supseteq \mathfrak{B}(H)$  and  $A\mathfrak{B}(H) = \mathfrak{D}$ , the identity  $(H + T)^* = H^* + T$ , which holds by virtue of the fact that  $T$  is bounded, implies that  $g$  is in  $\mathfrak{D}(H^*) = \mathfrak{D}(H^* + T)$  and that  $g^* = (H^* + T)g$ . Hence, since  $T$  is self-adjoint,  $\{g^*, g, h\}$  satisfies the above equation for all  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$  if and only if

<sup>†</sup> Stone, Theorem 2.22.

$$(f, H^*g) - (H^*f, g) + (Af, h) = 0$$

for all such  $\{f, H^*f\}$ . But the latter condition is satisfied if and only if  $h = WAg$ . Thus, in view of the remark following Definition 1.1 we conclude that  $C$  is a reduction operator for  $H^* + T$ .

To reveal more clearly the significance of Definition 1.1, we give now some concrete examples of reduction operators.

EXAMPLE 1.  $\mathfrak{S}$  is the space  $\mathfrak{L}_2(a, b)$  where  $a \leq x \leq b$  is a finite interval of the real axis.  $H^*$  is the transformation which takes  $f$  into  $if'$  and has for its domain  $\mathfrak{D}^*$  the set of all elements  $f$  of  $\mathfrak{L}_2(a, b)$  which are absolutely continuous on  $a \leq x \leq b$  and such that  $\int_a^b |f'|^2 dx < \infty$ . The operator  $H$  is the contraction of  $H^*$  whose domain  $\mathfrak{D}$  is the set of all elements in  $\mathfrak{D}^*$  which vanish at  $a$  and  $b$ . The space  $\mathfrak{M}$  is a two-dimensional unitary space, and  $A$  is the transformation which takes  $\{f, H^*f\}$  into the point of  $\mathfrak{M}$  with coordinates  $(f(b), f(a))$ ;  $W$  is the transformation which takes the point  $(c, d)$  in  $\mathfrak{M}$  into  $(ic, -id)$ . Here the formula (1.1) is the formula of Lagrange,

$$\int_a^b f \overline{ig'} dx - \int_a^b if' \overline{g} dx + f(b) \overline{ig(b)} - f(a) \overline{ig(a)} = 0.$$

EXAMPLE 2.  $\mathfrak{S}$  is the same as in Example 1; the domain  $\mathfrak{D}^*$  of  $H^*$  is the set of all elements  $f$  of  $\mathfrak{L}_2(a, b)$  such that  $f$  and  $f'$  are absolutely continuous on  $a \leq x \leq b$  and  $\int_a^b |f''|^2 dx$  is finite;  $H^*f = f''$ . The domain  $\mathfrak{D}$  of  $H$  consists of those and only those elements  $g$  of  $\mathfrak{D}^*$  such that  $g(a) = g(b) = g'(a) = g'(b) = 0$ ;  $Hg = g''$ . The space  $\mathfrak{M}$  is a four-dimensional unitary space;

$$A\{f, H^*f\} = \{f(b), f(a), f'(a), -f'(b)\}.$$

The transformation  $W$  takes  $(h, k, l, m)$  into  $(m, l, -k, -h)$  and the formula (1.1) is

$$\int_a^b f \overline{g'''} dx - \int_a^b \overline{g} f''' dx - f(b) \overline{g'(b)} + f(a) \overline{g'(a)} - f'(a) \overline{g(a)} + f'(b) \overline{g(b)} = 0.$$

EXAMPLE 3.  $\mathfrak{S}$  is the space  $\mathfrak{L}_2(E)$  where  $E$  is the set in the  $(x, y)$ -plane bounded by the lines  $x = a, x = b, y = c, y = d, a < b, c < d$ . The domain  $\mathfrak{D}^*$  of  $H^*$  is the set of all elements  $f(x, y)$  of  $\mathfrak{L}_2(E)$  which are absolutely continuous on  $a \leq x \leq b$  for almost all  $y$  on  $c < y < d$  and for which  $\int_E |f_x|^2 dE$  is finite;  $H^*f = if_x$ . The domain  $\mathfrak{D}$  of  $H$  is the set of all elements  $g$  of  $\mathfrak{D}^*$  such that  $\lim_{x \rightarrow a} g(x, y) = \lim_{x \rightarrow b} g(x, y) = 0$  for almost all  $y$  on  $c < y < d$ ;  $Hg = ig_x$ . The space  $\mathfrak{M}$  is the space  $\mathfrak{L}_2(c, d) \oplus \mathfrak{L}_2(c, d)$ ;

$$Af = \{f(b, y), f(a, y)\}; \quad W\{h(y), k(y)\} = -i\{h(y), -k(y)\}.$$

Here the formula (1.1) is

$$\int_E f \bar{ig}_z dE - \int_E if_x \bar{g} dE - \int_c^d f(b, y) \bar{ig}(b, y) dy + \int_c^d f(a, y) \bar{ig}(a, y) dy = 0,$$

and is valid for all  $f$  and  $g$  in  $\mathfrak{D}^*$ .

EXAMPLE 4.  $\mathfrak{E}$  is the space  $\mathfrak{E}_2(E)$  where  $E$  is the interior of the unit circle in the  $(x, y)$ -plane. The domain  $\mathfrak{D}^*$  of  $H^*$  is the set of all elements  $f$  of  $\mathfrak{E}_2(E)$  with the following properties:

(a)  $f, f_x$  are absolutely continuous in  $x$  on the closed intervals  $a \leq x \leq b$ ,  $y = k$  interior to  $E$  for almost all values of  $k$  on  $-1 < y < 1$ ;  $f, f_y$  are absolutely continuous in  $y$  on the closed intervals  $c \leq y \leq d$ ,  $x = h$ , interior to  $E$  for almost all values of  $h$  on  $-1 < x < 1$ ;

(b) The integrals

$$\int_S |f_x|^2 dS, \quad \int_S |f_y|^2 dS, \quad \int_S |f_{xx}|^2 dS, \quad \int_S |f_{yy}|^2 dS$$

exist on every closed set  $S$  interior to  $E$ ;

(c)  $-\nabla^2 f = -f_{xx} - f_{yy}$  is an element of  $\mathfrak{E}_2(E)$ .

$H^*$  is the transformation with domain  $\mathfrak{D}^*$  which takes  $f$  into  $-\nabla^2 f$ . The domain  $\mathfrak{D}$  of  $H$  is the subset of elements  $g$  of  $\mathfrak{D}^*$  such that  $g = g_x = g_y = 0$  almost everywhere on the boundary  $C$  of  $E$ . The space  $\mathfrak{M}$  is  $\mathfrak{E}_2(C) \oplus \mathfrak{E}_2(C)$ . The domain of  $A$  is the set of all elements  $\{f, H^*f\}$  of  $\mathfrak{B}(H^*)$  for which  $f(s)$  and  $\partial f / \partial n = -f_x y'(s) + f_y x'(s)$ , where  $s$  denotes arc length on  $C$ , are elements of  $\mathfrak{E}_2(C)$ ;

$$Af = \{f(s), -\partial f / \partial n\}.$$

$W$  is the transformation which takes  $\{h(s), k(s)\}$  into  $\{k(s), -h(s)\}$ . The formula (1.1) is here the familiar identity of Green,

$$-\int_E f \nabla^2 \bar{g} dE + \int_E \bar{g} \nabla^2 f dE - \int_C f \partial \bar{g} / \partial n \cdot ds + \int_C \partial f / \partial n \cdot \bar{g} ds = 0.$$

Here the domain of  $A$  is not identically  $\mathfrak{B}(H^*)$  as it is in each of the first three examples.

In later papers we shall deal in some detail with applications of the theory developed in the present memoir. Consequently we omit here proofs that the examples just given are valid illustrations of the situation described in Definition 1.1.

By means of Theorem 1.3, further examples are readily constructed from the ones just given. In particular; in Theorem 1.3 if  $H^*$  is a differential operator and  $T$  an integral operator, the sum  $H^* + T$  is an integro-differential oper-

ator of a type previously studied from various points of view by many writers.

Before proceeding, we point out that it is possible to define a reduction operator  $A$  on the graph of the adjoint of an arbitrary symmetric transformation  $H$ . A proof of this fact is given later (Theorem 2.9).

**2. Linear boundary conditions.** We introduce now a general definition of linear boundary conditions which describes, in particular, all the linear self-adjoint boundary conditions ordinarily considered in connection with the differential operators of Examples 1-4.

**DEFINITION 1.2.** Let  $A$  be a reduction operator for  $H^*$  with range-space  $\mathfrak{M}$ . Let  $\mathfrak{N}$  be a linear manifold in  $\mathfrak{M}$ , and let  $\mathfrak{D}(\mathfrak{N})$  be the set of elements  $f$  in the domain of  $H^*$  such that  $Af$  is defined and belongs to  $\mathfrak{N}$ . Then  $H(\mathfrak{N})$  denotes the contraction of  $H^*$  with domain  $\mathfrak{D}(\mathfrak{N})$  and  $Af \in \mathfrak{N}$  is called the boundary condition defining  $H(\mathfrak{N})$ . If the equation  $Af = h$  does not have a solution  $f$  for a dense set of elements  $h$  in  $\mathfrak{N}$ , the boundary condition is said to be degenerate; otherwise it is said to be nondegenerate.

**THEOREM 1.4.** The transformation  $H(\mathfrak{N})$  of Definition 1.2 is a linear extension of  $H$  and  $H^* \supseteq H^*(\mathfrak{N}) \supseteq H(\mathfrak{M} \ominus W\mathfrak{N})$ , where  $W$  has the same meaning as in Definition 1.1. If  $T$  is an arbitrary linear extension of  $H$  such that  $\mathfrak{B}(T) \subseteq \mathfrak{D}(A)$ , then  $\mathfrak{N} = A\mathfrak{B}(T)$  is a linear manifold in  $\mathfrak{M}$  and  $T \equiv H(\mathfrak{N})$ .

Since  $\mathfrak{N}$  is a linear manifold and  $A$  a linear transformation, the condition  $Af \in \mathfrak{N}$  clearly describes a linear manifold of elements  $f$  of  $\mathfrak{D}(H^*)$ . Thus, since  $H^*$  is linear,  $H(\mathfrak{N})$  is also. Furthermore, since  $\mathfrak{N}$  is linear, it contains the element 0 of  $\mathfrak{M}$ . Therefore, by Theorem 1.1,  $H(\mathfrak{N}) \supseteq H$ , and this implies  $H^* \supseteq H^*(\mathfrak{N})$ . The relation  $H^*(\mathfrak{N}) \supseteq H(\mathfrak{M} \ominus W\mathfrak{N})$  is an immediate consequence of Definition 1.2 and the identity (1.2).

If  $\mathfrak{B}(T) \subseteq \mathfrak{D}(A)$ , it follows at once, from the fact that  $\mathfrak{B}(T)$  is a linear manifold and  $A$  a linear transformation, that  $\mathfrak{N} = A\mathfrak{B}(T)$  is a linear manifold. The relation  $T \equiv H(\mathfrak{N})$  is a consequence of Definition 1.2 and the hypothesis  $\mathfrak{B}(T) \subseteq \mathfrak{D}(A)$ .

**3. The fundamental problem.** We are primarily concerned here with those transformations  $H(\mathfrak{N})$  which are symmetric. In order to isolate the boundary conditions  $Af \in \mathfrak{N}$  which define such extensions of  $H$ , we introduce the following definition:

**DEFINITION 1.3.** If  $\mathfrak{M}$  is a unitary or Hilbert space and  $W$  is a unitary transformation in  $\mathfrak{M}$  such that  $W^2 + I \equiv 0$ , a linear manifold  $\mathfrak{N}$  in  $\mathfrak{M}$  is said to be  $W$ -symmetric if  $W\mathfrak{N} \subseteq \mathfrak{M} \ominus \mathfrak{N}$ .

**THEOREM 1.5.** If  $\mathfrak{N}$  is a linear  $W$ -symmetric manifold, the transformation  $H(\mathfrak{N})$  of Definition 1.2 is a linear symmetric extension of  $H$ . If  $S$  is a linear

symmetric extension of  $H$  such that  $\mathfrak{B}(S) \subseteq \mathfrak{D}(A)$ , then  $A\mathfrak{B}(S)$  is a linear  $W$ -symmetric manifold in  $\mathfrak{M}$ .

Theorem 1.5 follows from Theorem 1.4 and the identity (1.2).

We can now state with some precision the twofold problem whose solution is our primary object.

**PROBLEM.** (1) To determine conditions on  $\mathfrak{N}$  necessary and sufficient for  $H(\mathfrak{N})$  to be maximal symmetric and conditions on  $\mathfrak{N}$  necessary and sufficient for  $H(\mathfrak{N})$  to be self-adjoint; (2) if  $\mathfrak{D}(A) \neq \mathfrak{B}(H^*)$ , to determine conditions on  $\mathfrak{N}$  necessary and sufficient for  $\tilde{H}(\mathfrak{N})$  to be maximal symmetric and conditions necessary and sufficient for  $\tilde{H}(\mathfrak{N})$  to be self-adjoint.

We leave to the reader the interpretation in terms of Examples 1–4, and in terms of others which he may construct, of the concepts introduced in this section and the one preceding.†

## CHAPTER II. $W$ -SYMMETRIC MANIFOLDS

**1. Isometric transformations.** The present chapter is devoted almost entirely to an analysis of the concept of  $W$ -symmetry introduced in Definition 1.3. This analysis is based on a simple correspondence between  $W$ -symmetric manifolds and isometric transformations which is immediately suggested by a well known correspondence between symmetric and isometric transformations,‡ and which has previously been described by O. Teichmüller.§ Before proceeding to the discussion of this correspondence, we state in a form adapted to our special purposes certain facts concerning isometric transformations.

**DEFINITION 2.1.** Let  $V$  be a closed isometric transformation from a space  $\mathfrak{S}_1$  to a space  $\mathfrak{S}_2$  where each of the spaces  $\mathfrak{S}_1, \mathfrak{S}_2$  is either a unitary space or a Hilbert space. Let  $m$  and  $n$  be, respectively, the dimension numbers of the manifolds  $\mathfrak{S}_1 \ominus \mathfrak{D}(V)$  and  $\mathfrak{S}_2 \ominus \mathfrak{R}(V)$ . Then the number pair  $(m, n)$  is called the  $(\mathfrak{S}_1, \mathfrak{S}_2)$ -deficiency index of  $V$ . If  $V$  is a non-closed isometric transformation from  $\mathfrak{S}_1$  to

† In this connection, the following material will be found suggestive: chap. 10, §§2, 3 of the book of Stone, especially Theorems 10.7, 10.16, and 10.18; the paper of I. Halperin, *Closures and adjoints of linear differential operators*, *Annals of Mathematics*, (2), vol. 38 (1937), pp. 880–919; especially pp. 883–891; the writer's abstract 43-3-114, *Bulletin of the American Mathematical Society*.

‡ J. von Neumann, *Mathematische Annalen*, vol. 102 (1929), pp. 49–131, especially pp. 80–91; or Stone, chap. 9.

§ *Operatoren im Wachschen Raum*, *Journal für die reine und angewandte Mathematik* (Crelle), vol. 174 (1935), pp. 73–124; especially pp. 99–107. Some of the theorems of the present chapter are only slight variations of results stated by Teichmüller. However, since his analysis does not lend itself readily to our purposes and would, in any case, have to be considerably supplemented at several points, it appeared to us desirable to carry through a complete independent treatment.

$\mathfrak{H}_2$ , the  $(\mathfrak{H}_1, \mathfrak{H}_2)$ -deficiency index of  $\tilde{V}$  is also said to be the  $(\mathfrak{H}_1, \mathfrak{H}_2)$ -deficiency index of  $V$ . If  $V$  has either domain identically  $\mathfrak{H}_1$  or range identically  $\mathfrak{H}_2$ ,  $V$  is said to be a maximal isometric transformation from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$ ; if both conditions are satisfied,  $V$  is said to be a unitary transformation from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$ .

**THEOREM 2.1.** Let  $\mathfrak{H}_1$  and  $\mathfrak{H}_2$  be separable complex euclidean spaces with dimension numbers  $j$  and  $k$ , respectively. A necessary and sufficient condition that a closed isometric transformation from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$  be unitary (maximal isometric) is that its  $(\mathfrak{H}_1, \mathfrak{H}_2)$ -deficiency index be  $(0, 0)$  (either  $(p, 0)$  or  $(0, p)$ ).

If either  $j$  or  $k$  is zero, the class of all isometric transformations from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$  contains only the linear transformation whose domain is  $\mathfrak{D}$ . If both  $j$  and  $k$  are different from zero, the class of all maximal isometric transformations from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$  has the cardinal number of the continuum.

Let  $j$  and  $k$  be different from zero. If  $j = k < \aleph_0$ , every maximal isometric transformation from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$  is unitary. If  $j \neq k$ , the  $(\mathfrak{H}_1, \mathfrak{H}_2)$ -deficiency index of every maximal isometric transformation from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$  is  $(j - k, 0)$  or  $(0, k - j)$  according as  $j$  is greater or less than  $k$ . If  $j$  is greater (less) than  $k$ , and  $\mathfrak{M}$  is an arbitrary closed linear manifold in  $\mathfrak{H}_1$  ( $\mathfrak{H}_2$ ) such that  $\mathfrak{H}_1 \ominus \mathfrak{M}$  ( $\mathfrak{H}_2 \ominus \mathfrak{M}$ ) has the dimension number  $k$  ( $j$ ), then every closed isometric transformation with domain  $\mathfrak{H}_1 \ominus \mathfrak{M}$  and range  $\mathfrak{H}_2$  (domain  $\mathfrak{H}_1$  and range  $\mathfrak{H}_2 \ominus \mathfrak{M}$ ) is contained in the class of all maximal isometric transformations from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$ . The class of all such transformations has the cardinal number of the continuum.

If  $j = k = \aleph_0$  and  $p$  is an arbitrary cardinal number on the range  $0 \leq p \leq \aleph_0$ , the set of all isometric transformations from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$  with the  $(\mathfrak{H}_1, \mathfrak{H}_2)$ -deficiency index  $(0, p)$  ( $(p, 0)$ ) has the cardinal number of the continuum. If  $\mathfrak{M}$  is an arbitrary closed linear manifold in  $\mathfrak{H}_1$  ( $\mathfrak{H}_2$ ) whose orthogonal complement has the dimension number  $\aleph_0$ , then every closed isometric transformation with domain  $\mathfrak{H}_1 \ominus \mathfrak{M}$  and range  $\mathfrak{H}_2$  (domain  $\mathfrak{H}_1$  and range  $\mathfrak{H}_2 \ominus \mathfrak{M}$ ) belongs to the class of all maximal isometric transformations from  $\mathfrak{H}_1$  to  $\mathfrak{H}_2$ . The class of all such transformations has the cardinal number of the continuum.

The assertions of the preceding theorem which are not obvious are either direct consequences of theorems of Stone (op. cit., chap. 2, §5) or are easily proved by methods similar to those used there. We therefore omit proof.

**2.  $W$ -symmetric manifolds and isometric transformations.** We state first the following definition:

**DEFINITION 2.2.** A  $W$ -symmetric manifold  $\mathfrak{N}$  in a space  $\mathfrak{M}$  is said to be maximal  $W$ -symmetric if it has no proper  $W$ -symmetric extension. It is said to be hypermaximal  $W$ -symmetric if  $W\mathfrak{N} = \mathfrak{M} \ominus \mathfrak{N}$ .

We come now to the fundamental theorem on  $W$ -symmetric manifolds.

**THEOREM 2.2.** *Let  $W$  be a unitary transformation in a unitary or Hilbert space  $\mathfrak{M}$ , and let  $W$  satisfy the identity  $W^2 + I \equiv 0$ . Let  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  be, respectively, the characteristic manifolds of  $W$  for the characteristic values  $+i$  and  $-i$ . Let  $\mathcal{U}_W$  be the class of all isometric transformations from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ , and let  $\mathcal{S}_W$  be the class of all linear  $W$ -symmetric manifolds  $\mathfrak{N}$  in  $\mathfrak{M}$ . Then there is a one-to-one correspondence between the classes  $\mathcal{U}_W$  and  $\mathcal{S}_W$  such that, if  $V$  and  $\mathfrak{N}$  correspond,  $\mathfrak{N}$  is the range of  $I - V$ . If  $V$  and  $\mathfrak{N}$  correspond, so also do  $\bar{V}$  and  $\bar{\mathfrak{N}}$ ;  $\mathfrak{N}$  is a closed linear manifold if and only if  $V$  is a closed transformation. The correspondence between  $\mathcal{U}_W$  and  $\mathcal{S}_W$  is an isomorphism with respect to the relation  $\subset$ . If  $\mathfrak{N}$  is a closed linear manifold in  $\mathcal{S}_W$  and  $V$  its correspondent in  $\mathcal{U}_W$ , then*

$$(2.1) \quad \mathfrak{M} \ominus \mathfrak{N} = W\mathfrak{N} + (\mathfrak{M}^+ \ominus \mathfrak{D}(V)) + (\mathfrak{M}^- \ominus \mathfrak{R}(V)), \dagger$$

*the three component manifolds in the right-hand member being mutually orthogonal. A manifold  $\mathfrak{N}$  in  $\mathcal{S}_W$  is maximal (hypermaximal)  $W$ -symmetric if and only if the correspondent  $V$  in  $\mathcal{U}_W$  is a maximal isometric (unitary) transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ .*

We observe first that  $\mathfrak{M}^- = \mathfrak{M} \ominus \mathfrak{M}^+$ . This follows from the unitary character of  $W$  and the easily established fact that the spectrum of  $W$  consists only of the two points  $+i$  and  $-i$ .

Consider now an arbitrary manifold  $\mathfrak{N}$  of the class  $\mathcal{S}_W$ . Let  $f$  be an arbitrary element of  $\mathfrak{N}$ . Then  $f$  has a unique resolution  $f = f^+ + f^-$ , where  $f^+$  is in  $\mathfrak{M}^+$ ,  $f^-$  in  $\mathfrak{M}^-$ . Furthermore,

$$(f, Wf) = (f^+ + f^-, if^+ - if^-) = -i(f^+, f^+) + i(f^-, f^-) = 0.$$

Thus  $|f^-| = |f^+|$ , and we can set  $f^- = -Vf^+$ , where  $V$  is an operator with domain the set of elements of  $\mathfrak{M}^+$  which are projections of elements of  $\mathfrak{N}$  and is uniquely defined at every point of its domain. Since  $\mathfrak{N}$  is a linear manifold and the projections with ranges  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  are linear transformations,  $V$  is a linear transformation. Since, as we have already seen,  $V$  preserves norm,  $V$  is isometric. It is obvious that  $V$  belongs to the class  $\mathcal{U}_W$  and that  $\mathfrak{N}$  is the range of  $I - V$ . On the other hand, if  $V$  is an arbitrary element of the class  $\mathcal{U}_W$ , we have, for all  $f$  and  $g$  in the domain of  $V$ ,

$$(f - Vf, W(g - Vg)) = (f - Vf, ig + iVg) = -i(f, g) + i(Vf, Vg) = 0.$$

Thus the range of  $I - V$  belongs to  $\mathcal{S}_W$ . Since it is evident that two isometric transformations  $V$  and  $V_1$  with domains in  $\mathfrak{M}^+$  and ranges in  $\mathfrak{M}^-$  cannot satisfy the relation  $\mathfrak{N}(I - V) = \mathfrak{N}(I - V_1)$  unless  $V \equiv V_1$ , the first assertion of the theorem is established.

$\dagger$  We use the ordinary plus sign here and throughout this paper to indicate the linear sum of manifolds.

We now turn to the second. If  $V$  is an arbitrary member of the class  $\mathcal{U}_W$  and  $\{f_n\}$  an arbitrary convergent sequence in the domain of  $V$ , the sequence  $\{f_n - Vf_n\}$  also converges, evidently to an element  $f - Vf$  of  $\mathfrak{N}$ . Furthermore, if  $\{f_n\}$  is an arbitrary convergent sequence in  $\mathfrak{N}$  and  $f_n = f_n^+ - Vf_n^+$ , the sequences  $\{f_n^+\}$  and  $\{Vf_n^+\}$  must converge separately since they belong to orthogonal linear manifolds. But then the first converges to an element  $f^+$  in the domain of  $\tilde{V}$ , the second to  $\tilde{V}f^+$ . Hence if  $\mathfrak{N}$  and  $V$  are in correspondence, so also are  $\tilde{\mathfrak{N}}$  and  $\tilde{V}$ . Since the correspondence is one-to-one, it follows that  $\mathfrak{N}$  is closed if and only if  $V$  is closed.

Purely on the basis of the definition of the correspondence between  $\mathcal{S}_W$  and  $\mathcal{U}_W$  it is readily verified that  $\mathfrak{N}_1 \subset \mathfrak{N}_2$  implies  $V_1 \subseteq V_2$  while  $V_1 \subset V_2$  implies  $\mathfrak{N}_1 \subseteq \mathfrak{N}_2$ , where, in either case,  $\mathfrak{N}_1$  and  $\mathfrak{N}_2$  correspond to  $V_1$  and  $V_2$ , respectively. Since the correspondence is one-to-one,  $V_1 = V_2$  if and only if  $\mathfrak{N}_1 = \mathfrak{N}_2$ . Hence  $\mathfrak{N}_1 \subset \mathfrak{N}_2$  implies  $V_1 \subset V_2$ , and conversely. Furthermore, since a maximal isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  has no proper maximal isometric extension from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ , the equivalence, when  $\mathfrak{N}$  and  $V$  correspond, of the statements " $\mathfrak{N}$  is maximal  $W$ -symmetric" and " $V$  is a maximal isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ " follows at once.

To establish the relation

$$\mathfrak{M} \ominus \mathfrak{N} = W\mathfrak{N} + (\mathfrak{M}^+ \ominus \mathfrak{D}(V)) + (\mathfrak{M}^- \ominus \mathfrak{R}(V)),$$

where  $\mathfrak{N}$  is a closed linear manifold in  $\mathcal{S}_W$ ,  $V$  the corresponding member of  $\mathcal{U}_W$ , we first observe that it is equivalent to the equation

$$\mathfrak{M} \ominus [(\mathfrak{M}^+ \ominus \mathfrak{D}(V)) + (\mathfrak{M}^- \ominus \mathfrak{R}(V))] = \mathfrak{N} + W\mathfrak{N}.$$

But, since  $\mathfrak{M}^- = \mathfrak{M} \ominus \mathfrak{M}^+$ , we have

$$\mathfrak{M} \ominus [(\mathfrak{M}^+ \ominus \mathfrak{D}(V)) + (\mathfrak{M}^- \ominus \mathfrak{R}(V))] = \mathfrak{D}(V) + \mathfrak{R}(V).$$

Hence we need only show that

$$\mathfrak{N} + W\mathfrak{N} = \mathfrak{D}(V) + \mathfrak{R}(V).$$

As every element of  $\mathfrak{N}$  can be written in the form  $f^+ - Vf^+$  and every element of  $W\mathfrak{N}$  in the form  $g^+ + Vg^+$ , where  $f^+$  and  $g^+$  belong to  $\mathfrak{D}(V)$ , we have immediately  $\mathfrak{N} + W\mathfrak{N} \subseteq \mathfrak{D}(V) + \mathfrak{R}(V)$ . On the other hand, if  $f^+$  is an arbitrary element of  $\mathfrak{D}(V)$ ,  $(f^+ - Vf^+)/2$  is in  $\mathfrak{N}$  while  $(f^+ + Vf^+)/2$  is in  $W\mathfrak{N}$ , whence we conclude that

$$f^+ = (f^+ - Vf^+)/2 + (f^+ + Vf^+)/2$$

belongs to  $\mathfrak{N} + W\mathfrak{N}$ . Similarly, if  $Vg^+$  is an arbitrary element of  $\mathfrak{R}(V)$ ,  $-(g^+ - Vg^+)/2$  and  $(g^+ + Vg^+)/2$  belong to the respective manifolds  $\mathfrak{N}$  and

$W\mathfrak{N}$ . Hence  $Vg^+$  is in  $\mathfrak{N} + W\mathfrak{N}$ . Thus  $\mathfrak{N} + W\mathfrak{N} = \mathfrak{D}(V) + \mathfrak{R}(V)$  as we wished to show.

That  $\mathfrak{N}$  is hypermaximal  $W$ -symmetric if and only if the corresponding member of  $\mathcal{U}_W$  is a unitary transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  is an immediate consequence of the result established in the preceding paragraph. The proof of the theorem is therefore complete.

Throughout the remainder of this chapter,  $W$ ,  $\mathfrak{M}$ ,  $\mathfrak{M}^+$ ,  $\mathfrak{M}^-$ ,  $\mathcal{U}_W$ ,  $\mathcal{S}_W$  have the same meanings as in Theorem 2.2. We shall hereafter write  $\mathfrak{N} = \mathfrak{N}(V)$  and  $V = V(\mathfrak{N})$  to indicate that  $\mathfrak{N}$  and  $V$  are corresponding members of  $\mathcal{S}_W$  and  $\mathcal{U}_W$ , respectively.

**THEOREM 2.3.** *Let  $\mathfrak{M}_1^+$ ,  $\mathfrak{M}_2^+$ ,  $\dots$ ,  $\mathfrak{M}_n^+$  be mutually orthogonal linear manifolds in  $\mathfrak{M}^+$ . Let  $V_1, V_2, \dots, V_n$  be transformations of the class  $\mathcal{U}_W$  with domains respectively  $\mathfrak{M}_1^+, \mathfrak{M}_2^+, \dots, \mathfrak{M}_n^+$ , and ranges mutually orthogonal. Then the manifolds  $\mathfrak{N}(V_1), \mathfrak{N}(V_2), \dots, \mathfrak{N}(V_n)$  are mutually orthogonal. If  $V$  is the linear transformation with domain  $\mathfrak{M}_1^+ + \mathfrak{M}_2^+ + \dots + \mathfrak{M}_n^+$  which is equal on  $\mathfrak{M}_k$  to  $V_k$ , ( $k = 1, 2, \dots, n$ ),  $V$  belongs to  $\mathcal{U}_W$  and  $\mathfrak{N}(V)$  is  $\mathfrak{N}(V_1) + \mathfrak{N}(V_2) + \dots + \mathfrak{N}(V_n)$ . Conversely, if  $\mathfrak{N}_1, \mathfrak{N}_2, \dots, \mathfrak{N}_n$  are mutually orthogonal linear manifolds in  $\mathcal{S}_W$ , the domains, and the ranges, of  $V(\mathfrak{N}_1), V(\mathfrak{N}_2), \dots, V(\mathfrak{N}_n)$  are mutually orthogonal. If  $\mathfrak{N} = \mathfrak{N}_1 + \mathfrak{N}_2 + \dots + \mathfrak{N}_n$ ,  $\mathfrak{N}$  belongs to  $\mathcal{S}_W$  and  $V(\mathfrak{N})$  has domain the linear manifold determined by the domains of  $V(\mathfrak{N}_1), V(\mathfrak{N}_2), \dots, V(\mathfrak{N}_n)$  and is equal to  $V(\mathfrak{N}_k)$  on the domain of the latter ( $k = 1, 2, \dots, n$ ).*

The truth of Theorem 2.3 follows from the definition of the correspondence of Theorem 2.2; we omit the details of proof.

**DEFINITION 2.3.** *The  $(\mathfrak{M}^+, \mathfrak{M}^-)$ -deficiency index of an isometric transformation  $V$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  is also called the  $W$ -deficiency index of  $V$ . If  $\mathfrak{N}$  is a linear  $W$ -symmetric manifold and  $V(\mathfrak{N})$  has the  $W$ -deficiency index  $(m, n)$ ,  $(m, n)$  is also said to be the  $W$ -deficiency index of  $\mathfrak{N}$ .*

**THEOREM 2.4.** *A necessary and sufficient condition that a closed linear manifold  $\mathfrak{N}$  of the class  $\mathcal{S}_W$  be maximal  $W$ -symmetric is that its  $W$ -deficiency index be either  $(m, 0)$  or  $(0, n)$ . A necessary and sufficient condition that  $\mathfrak{N}$  be hypermaximal  $W$ -symmetric is that its  $W$ -deficiency index be  $(0, 0)$ .*

The preceding theorem is merely a restatement, in terms of the terminology introduced in Definition 2.3 of the last two assertions of Theorem 2.2.

**THEOREM 2.5.** *Let  $j$  be the dimension number of  $\mathfrak{M}^+$ ,  $k$  the dimension number of  $\mathfrak{M}^-$ . Then, if either  $j$  or  $k$  is zero,  $\mathcal{S}_W$  contains only the manifold  $\mathfrak{D}$  in  $\mathfrak{M}$ . If neither  $j$  nor  $k$  is zero, the class of all maximal  $W$ -symmetric manifolds in  $\mathcal{S}_W$  has the cardinal number of the continuum.*

If  $j = k < \aleph_0$ , every maximal  $W$ -symmetric manifold in the class  $\mathcal{S}_W$  is hyper-maximal; that is, has the  $W$ -deficiency index  $(0, 0)$ . If  $j > k$  ( $k > j$ ) holds, every maximal  $W$ -symmetric manifold in  $\mathcal{S}_W$  has the  $W$ -deficiency index  $(j - k, 0)$  ( $(0, k - j)$ ). If  $j = k = \aleph_0$  and  $p$  is a cardinal number on the range  $0 \leq p \leq \aleph_0$ , the class of all maximal  $W$ -symmetric manifolds with the  $W$ -deficiency index  $(p, 0)$  ( $(0, p)$ ) has the cardinal number of the continuum.

If  $j = k = \aleph_0$ , or if  $j > k$  ( $k > j$ ) holds, and  $\mathcal{M}_1$  is an arbitrary closed linear manifold in  $\mathcal{M}^+$  ( $\mathcal{M}^-$ ) such that  $\mathcal{M}^+ \ominus \mathcal{M}_1$  ( $\mathcal{M}^- \ominus \mathcal{M}_1$ ) has the dimension number  $k(j)$ , the class of all isometric transformations of  $\mathcal{M}^+ \ominus \mathcal{M}_1$  into  $\mathcal{M}^-$  ( $\mathcal{M}^+$  into  $\mathcal{M}^- \ominus \mathcal{M}_1$ ) is in one-to-one correspondence according to Theorem 2.2 with a subclass of the class of all maximal  $W$ -symmetric manifolds in  $\mathcal{S}_W$ . Each member of the subclass of  $\mathcal{S}_W$  so defined has the  $W$ -deficiency index  $(p, 0)$  ( $(0, p)$ ), where  $p$  is the dimension number of  $\mathcal{M}_1$ .

Theorem 2.5 is for the most part a straightforward interpretation in terms of the class  $\mathcal{S}_W$  of the analysis of the maximal transformations in the class  $\mathcal{U}_W$  which is provided by Theorem 3.1. We leave the details of verification to the reader.

While it is evidently possible to elaborate extensively the preceding analysis and, in particular, to consider the problem of determining all of the maximal  $W$ -symmetric extensions of a given  $W$ -symmetric manifold, we refrain from such elaboration as being both unnecessary for the applications which we intend to make and also clearly suggested in outline by the analysis which we have already developed and by works of similar nature to which we referred at the beginning of this chapter.

**3. Illustrations and applications.** We pass instead to the consideration of two special types of transformation  $W$  which are of considerable importance and interest.

**THEOREM 2.6.** Let  $\mathcal{M} = \mathcal{S} \oplus \mathcal{S}$  where  $\mathcal{S}$  is a unitary or Hilbert space, and let  $W$  be the transformation which takes every vector  $\{f_1, f_2\}$  of  $\mathcal{S}$  into  $i\{f_1, -f_2\}$ . Then  $W$  is a unitary transformation in  $\mathcal{M}$  and  $W^2 + I = 0$ . The characteristic manifolds  $\mathcal{M}^+$  and  $\mathcal{M}^-$  of  $W$  for the characteristic values  $+i$  and  $-i$  are  $\mathcal{S} + \mathcal{D}$  and  $\mathcal{D} + \mathcal{S}$ , respectively. The class  $\mathcal{U}_W$  is in one-to-one correspondence with the class of all isometric transformations  $X$  in  $\mathcal{S}$ ;  $V$  in  $\mathcal{U}_W$  corresponds to  $X$  if and only if  $\mathcal{D}(V) = \mathcal{D}(X) + \mathcal{D}$  and  $V\{f, 0\} = \{0, Xf\}$  for all  $f$  in the domain of  $X$ .  $V$  is a maximal isometric (unitary) transformation from  $\mathcal{M}^+$  to  $\mathcal{M}^-$  if and only if  $X$  is maximal isometric (unitary) in  $\mathcal{S}$ . If  $V$  and  $X$  correspond,  $\mathcal{R}(V)$  is the graph of  $-X$ .

The statements regarding  $W$  are evident. The remaining assertions of the

theorem are all proved by simple recourse to the definitions of the various terms and symbols involved.

**THEOREM 2.7.** *Let  $\mathfrak{M} = \mathfrak{S} \oplus \mathfrak{S}$  where  $\mathfrak{S}$  is a unitary or Hilbert space. Let  $W$  be the transformation with domain  $\mathfrak{M}$  which takes  $\{f, g\}$  into  $\{g, -f\}$ . Then  $W$  is unitary and  $W^2 + I = 0$ . The manifolds  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  in  $\mathfrak{M}$  are, respectively, the set of vectors of the form  $\{h, ih\}$  and the set of vectors of the form  $\{h, -ih\}$ . There is a one-to-one correspondence between the class  $\mathcal{U}_W$  and the class  $\mathcal{X}$  of all isometric transformations  $X$  in  $\mathfrak{S}$ ;  $V$  and  $X$  correspond provided  $\{h, ih\}$  is in the domain of  $V$  when and only when  $h$  is in the domain of  $X$  and  $V\{h, ih\} = \{Xh, -iXh\}$ .*

*The class  $\mathcal{S}_W$  contains a subclass of manifolds each of which is the graph  $\mathfrak{B}(H)$  of a symmetric transformation  $H$  in  $\mathfrak{S}$ ; if  $H$  is an arbitrary linear symmetric transformation in  $\mathfrak{S}$ ,  $\mathfrak{B}(H)$  belongs to  $\mathcal{S}_W$ . If  $\mathfrak{B}(H) = \mathfrak{R}(V)$  and  $V$  corresponds to the member of  $\mathcal{X}$  according to the laws stated above,  $H \equiv i(I + X)(I - X)^{-1}$ .  $H$  is maximal symmetric (self-adjoint) if and only if  $\mathfrak{B}(H)$  is maximal (hypermaximal)  $W$ -symmetric.*

Again the statements concerning  $W$  are evident from inspection. In verification of the correspondence between  $\mathcal{U}_W$  and  $\mathcal{X}$ , we observe first that every isometric transformation  $X$  in  $\mathcal{X}$  clearly generates a member  $V$  of the class  $\mathcal{U}_W$  according to the following rules:

- (1)  $\mathfrak{D}(V)$  contains  $\{h, ih\}$  if and only if the domain of  $X$  contains  $h$ ;
- (2)  $V\{h, ih\} = \{Xh, -iXh\}$ .

On the other hand, if  $V$  belongs to  $\mathcal{U}_W$ , the equality  $2|h|^2 = 2|k|^2$  which is valid whenever  $\{k, -ik\} = V\{h, ih\}$  and the linear character of  $V$  imply that  $V$  generates a member  $X$  of  $\mathcal{X}$  such that (1) and (2) are satisfied.

If  $H$  is a linear symmetric transformation in  $\mathfrak{S}$ , then  $W\mathfrak{B}(H)$  is the set of vectors  $\{Hf, -f\}$  in  $\mathfrak{S} \oplus \mathfrak{S}$ ; hence the symmetry of  $H$  implies that  $\mathfrak{B}(H)$  is in  $\mathcal{S}_W$ . If  $\mathfrak{B}(H) = \mathfrak{R}(V)$  and  $V$  corresponds to  $X$  in  $\mathcal{X}$ , the domain of  $H$  is the range of  $I - X$  and  $H(I - X) \equiv i(I + X)$ . The fact that  $I - X$  has range dense in  $\mathfrak{S}$  is readily shown to imply that it has no zeros and we thus have  $H \equiv i(I + X)(I - X)^{-1}$ . The statements " $H$  is maximal symmetric" and " $\mathfrak{B}(H)$  is maximal  $W$ -symmetric" are obviously equivalent by definition as also are the statements " $H$  is self-adjoint" and " $\mathfrak{B}(H)$  is hypermaximal  $W$ -symmetric."

Further application of the theory of  $W$ -symmetric manifolds which we have developed, and of the more detailed theory which we have suggested to the study of the graphs of symmetric transformations would clearly yield new proofs of many of the known facts in the familiar theory of the connection between isometric and symmetric transformations. We prefer, however,

to postpone consideration of this theory to a later point where we are led to it again from a different direction. Here we shall prove only a variation of one known theorem which is intimately connected with the previously stated fact that a reduction operator  $A$  can be defined on the graph of the adjoint of any symmetric transformation.

**THEOREM 2.8.** *Let  $H$  be a closed linear symmetric transformation in Hilbert space  $\mathfrak{H}$ , and let  $\mathfrak{D}^+$  and  $\mathfrak{D}^-$  be the characteristic manifolds of  $H^*$  for the characteristic values  $+i$  and  $-i$ , respectively. Then  $\mathfrak{B}(H^*) \ominus \mathfrak{B}(H)$  is the set of elements of  $\mathfrak{B}(H^*)$  which can be written in the form  $\{f^+ + f^-, if^+ - if^-\}$ , where  $f^+$  is in  $\mathfrak{D}^+$ ,  $f^-$  in  $\mathfrak{D}^-$ .†*

Let  $W$ ,  $\mathfrak{M}^+$ , and  $\mathfrak{M}^-$  have meanings the same as in Theorem 2.7, and let  $V$  be the isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  corresponding to  $\mathfrak{B}(H)$  in accordance with that theorem. Since  $H$  is closed and linear,  $\mathfrak{B}(H)$  is a closed linear manifold, and  $V$  is closed also. Moreover, according to equation (2.1),  $(\mathfrak{S} \oplus \mathfrak{S}) \ominus \mathfrak{B}(H)$  is the manifold

$$W\mathfrak{B}(H) \oplus (\mathfrak{M}^+ \ominus \mathfrak{D}(V)) \oplus (\mathfrak{M}^- \ominus \mathfrak{R}(V)).$$

But, by definition,

$$(\mathfrak{S} \oplus \mathfrak{S}) \ominus \mathfrak{B}(H) = W\mathfrak{B}(H^*).$$

Thus we have

$$W\mathfrak{B}(H) \oplus (\mathfrak{M}^+ \ominus \mathfrak{D}(V)) \oplus (\mathfrak{M}^- \ominus \mathfrak{R}(V)) = W\mathfrak{B}(H^*),$$

and, since the three components on the left are mutually orthogonal, this is equivalent to the equation

$$W\mathfrak{B}(H^*) \ominus W\mathfrak{B}(H) = (\mathfrak{M}^+ \ominus \mathfrak{D}(V)) \oplus (\mathfrak{M}^- \ominus \mathfrak{R}(V)).$$

But  $W$  is unitary, equal to  $iI$  on  $\mathfrak{M}^+$ , and to  $-iI$  on  $\mathfrak{M}^-$ , so that the latter equation is equivalent to

$$\mathfrak{B}(H^*) \ominus \mathfrak{B}(H) = (\mathfrak{M}^+ \ominus \mathfrak{D}(V)) \oplus (\mathfrak{M}^- \ominus \mathfrak{R}(V)).$$

Hence to complete the proof it is necessary only to show that  $(\mathfrak{M}^+ \ominus \mathfrak{D}(V))$  consists of all vectors  $\{f^+, if^+\}$  such that  $f^+$  is in  $\mathfrak{D}^+$  and  $(\mathfrak{M}^- \ominus \mathfrak{R}(V))$  of all vectors  $\{f^-, if^-\}$  such that  $f^-$  is in  $\mathfrak{D}^-$ .

Consider first an arbitrary element  $f^+$  of  $\mathfrak{D}^+$ . Then  $H^*f^+ = if^+$ , and

$$(g, f^+) + (Hg, if^+) = i(g, if^+) - i(Hg, f^+) = 0$$

for all  $g$  in  $\mathfrak{D}(H)$ ; and  $\{f^+, if^+\}$  is perpendicular to  $\mathfrak{B}(H)$ . But  $\mathfrak{B}(H) = \mathfrak{R}(I - V)$  and  $\mathfrak{R}(V)$  is in  $\mathfrak{M}^- = (\mathfrak{S} \oplus \mathfrak{S}) \ominus \mathfrak{M}^+$ . Therefore, since  $(f^+, h) + (if^+, -ih) = 0$  for

† Compare Stone, Theorem 9.4, and Teichmüller, loc. cit., pp. 104-105.

all  $h$  in  $\mathfrak{S}$ ,  $\{f^+, if^+\}$  is in  $\mathfrak{M}^+ \ominus \mathfrak{D}(V)$ . On the other hand, if  $\{f^+, if^+\}$  is in  $\mathfrak{M}^+ \ominus \mathfrak{D}(V)$ , it is perpendicular to  $\mathfrak{B}(H)$  and we have

$$(g, if^+) - (Hg, f^+) = -i(g, f^+) + (Hg, if^+) = 0$$

for all  $g$  in  $\mathfrak{D}(H)$ , whence it follows that  $if^+ = H^*f^+$ .

A similar argument establishes the analogous relation between  $\mathfrak{M}^- \ominus \mathfrak{R}(V)$  and  $\mathfrak{D}^-$ , and completes the proof of the theorem.

**THEOREM 2.9.** *Let  $A$  be the projection in  $\mathfrak{B}(H^*)$  with range  $\mathfrak{B}_1(H^*) = \mathfrak{B}(H^*) \ominus \mathfrak{B}(H)$ . Then  $A$  is a reduction operator for  $H^*$ , with range-space  $\mathfrak{B}_1(H^*)$ . The transformation  $W$  associated with  $A$  by Definition 1.1 is equal on its domain to the operator in  $\mathfrak{S} \oplus \mathfrak{S}$  which takes  $\{f, g\}$  into  $\{g, -f\}$ , where  $f$  and  $g$  are both in  $\mathfrak{S}$ .*

It is necessary for the proof of the theorem only to determine

$$((\mathfrak{S} \oplus \mathfrak{S}) \oplus \mathfrak{B}_1(H^*)) \ominus \mathfrak{B}(A).$$

Letting  $W_1$  have the same meaning as  $W$  in Theorem 2.7, we have  $\mathfrak{S} \oplus \mathfrak{S} = \mathfrak{B}(H^*) \oplus W_1\mathfrak{B}(H)$ . Hence

$$((\mathfrak{S} \oplus \mathfrak{S}) \oplus \mathfrak{B}_1(H^*)) \ominus \mathfrak{B}(A) = ((\mathfrak{B}(H^*) \oplus W_1\mathfrak{B}(H)) \oplus \mathfrak{B}_1(H^*)) \ominus \mathfrak{B}(A),$$

and, since  $W_1\mathfrak{B}(H) + \mathfrak{D}$  is obviously orthogonal to  $\mathfrak{B}(A)$  in  $(\mathfrak{S} \oplus \mathfrak{S}) \oplus \mathfrak{B}_1(H^*)$ , we have now to determine

$$(\mathfrak{B}(H^*) \oplus \mathfrak{B}_1(H^*)) \ominus \mathfrak{B}(A).$$

But, considered as a transformation from  $\mathfrak{B}(H^*)$  to  $\mathfrak{B}_1(H^*)$ ,  $A$  has an adjoint from  $\mathfrak{B}_1(H^*)$  to  $\mathfrak{B}(H^*)$ , which is clearly equal on its domain to the identity. Therefore, in accordance with Theorem 2.8,

$$((\mathfrak{S} \oplus \mathfrak{S}) \oplus \mathfrak{B}_1(H^*)) \ominus \mathfrak{B}(A)$$

consists of those and only those elements of  $(\mathfrak{S} \oplus \mathfrak{S}) \oplus \mathfrak{B}_1(H^*)$  which are of the form

$$\{Hf + if^+ - if^-, -f - f^+ - f^-, \{-if^+ + if^-, f^+ + f^-\}\}.$$

Since, when we again take account of Theorem 2.8, these elements are revealed as precisely those elements of  $(\mathfrak{S} \oplus \mathfrak{S}) \oplus \mathfrak{B}_1(H^*)$  which are of the form  $\{H^*g, -g, WAg\}$ , where  $W$  has the character stated in the theorem, the proof is complete.

Theorem 2.6 gives complete information about the structure of the manifolds  $\mathfrak{R}$  of the class  $\mathfrak{S}_W$  for the special operator  $W$  there considered. Theorem 2.7, however, provides this information only in the special case that  $\mathfrak{R}$  is the graph of a symmetric transformation. It is desirable, therefore, that we dis-

cuss more fully the general case arising under Theorem 2.7. An analysis adequate for our purposes is provided by the theorem which we prove next. Following von Neumann, we shall call a transformation  $H$  in a space  $\mathfrak{S}$  Hermitian if the equation  $(f, Hg) - (Hf, g) = 0$  is satisfied for all  $f$  and  $g$  in the domain of  $H$ .† Thus a Hermitian transformation is symmetric if its domain determines  $\mathfrak{S}$ . We emphasize that  $\mathfrak{S}$  may have a finite dimension number and also that we do not admit, as von Neumann does in the work just referred to, many-valued operators  $H$ .

**THEOREM 2.10.** *Let  $\mathfrak{M}$  and  $W$  have the same meanings as in Theorem 2.7, and let  $\mathfrak{R}$  be an arbitrary closed linear manifold in the class  $\mathfrak{S}_W$ . Let  $X$  be the member of the class  $\mathfrak{X}$  corresponding to  $V(\mathfrak{R})$  in accordance with Theorem 2.7. Then  $X$  is a closed transformation. Let  $\mathfrak{D}_1(X)$  be the characteristic manifold of  $X$  for the characteristic value 1, and let  $\mathfrak{D}_0(X) = \mathfrak{D}(X) \ominus \mathfrak{D}_1(X)$ . Let  $X_0$  be the contraction of  $X$  with domain  $\mathfrak{D}_0(X)$ , and let  $\mathfrak{S}_0$  be the closed linear manifold determined by the linear sum of the domain and range of  $X_0$ . Then  $\mathfrak{S}_0$  is in  $\mathfrak{S} \ominus \mathfrak{D}_1(X)$ ,  $(I - X_0)^{-1}$  exists, and the transformation*

$$H \equiv i(I + X_0)(I - X_0)^{-1}$$

*is a closed linear Hermitian transformation in  $\mathfrak{S}_0$ . If  $\{f, g\}$  is an arbitrary element of  $\mathfrak{R}$ ,  $f$  is in the domain of  $H$  and  $g = Hf + h$ , where  $h$  belongs to  $\mathfrak{D}_1(X)$ . The resolution of  $\{f, g\}$  so provided is unique.  $\mathfrak{R}$  is maximal (hypermaximal)  $W$ -symmetric if and only if  $\mathfrak{S}_0 = \mathfrak{S} \ominus \mathfrak{D}_1(X)$  and  $H$  is maximal symmetric (self-adjoint) in  $\mathfrak{S}_0$ .*

That  $X$  is closed follows at once from the fact that  $\mathfrak{R}$  is closed, when we take account of the relation between  $X$  and  $V(\mathfrak{R})$  described in Theorem 2.7.

By definition,  $\mathfrak{S}_0 = \mathfrak{D}(X_0) \oplus \mathfrak{R}(X_0)$ . Since, also by definition,  $\mathfrak{D}(X_0) = \mathfrak{D}(X) \ominus \mathfrak{D}_1(X)$ , it follows that  $\mathfrak{D}(X_0) \subseteq \mathfrak{S} \ominus \mathfrak{D}_1(X)$ . But  $X_0 \subseteq X$  and  $X$  is isometric. Therefore, as  $X$  is defined throughout  $\mathfrak{D}_1(X)$  and leaves it invariant, we must have  $\mathfrak{R}(X_0) \subseteq \mathfrak{S} \ominus \mathfrak{D}_1(X)$ . Consequently  $\mathfrak{S}_0 \subseteq \mathfrak{S} \ominus \mathfrak{D}_1(X)$ .

Since  $(I - X_0)f = 0$  implies  $(I - X)f = 0$  which implies in turn that  $f$  belongs to  $\mathfrak{D}_1(X)$ , it follows that  $(I - X_0)^{-1}$  exists. That the transformation  $H$  of the theorem is a closed linear Hermitian transformation in the space  $\mathfrak{S}_0$  is readily verified on the basis of its definition in terms of  $X_0$ .

If  $\{f, g\}$  is an arbitrary element of  $\mathfrak{R}$ , then, according to Theorem 2.7,  $f = k - Xk$ ,  $g = i(k + Xk)$ , where  $k$  is an element of  $\mathfrak{D}(X)$ . Let  $k = k_1 + k_2$ , where  $k_1$  belongs to  $\mathfrak{D}_0(X)$ ,  $k_2$  to  $\mathfrak{D}_1(X)$ . Then  $f = k_1 - Xk_1$ ,  $g = i(k_1 + X_0k_1) + 2ik_2$ . Thus  $g = Hf + h$ , where  $h = 2ik_2$ . That this resolution of  $\{f, g\}$  is unique is an

† *Functional Operators* (mimeographed notes of lectures given at The Institute for Advanced Study, 1933-1935), p. 2-46.

immediate consequence of the uniqueness of the resolutions  $k = k_1 + k_2$  for  $k$ , and

$$\{f, g\} = \{k - Xk, i(k + Xk)\}$$

for  $\{f, g\}$ .

Now let us suppose that  $\mathfrak{N}$  is maximal  $W$ -symmetric. Then, by Theorem 2.7, either the domain or the range of  $X$  is identically  $\mathfrak{S}$ . Consequently, either the domain or the range of  $X_0$  is identically  $\mathfrak{S}_0$  and  $\mathfrak{S}_0 = \mathfrak{S} \ominus \mathfrak{D}_1(X)$ . Furthermore, the range of  $I - X_0$  is dense in  $\mathfrak{S}_0$ . For let the equation  $(f, g - X_0g) = 0$  be satisfied for all  $g$  in the domain of  $X_0$  and some element  $f$  of  $\mathfrak{S}_0$ . Then  $(f, g) - (f, X_0g) = 0$ . Hence, if  $X_0$  has domain  $\mathfrak{S}_0$ , we have  $f = X_0^*f$ . But it is readily shown that  $X_0^*$  is equal to  $X_0^{-1}$  on  $\mathfrak{R}(X_0)$  and equal to 0 on  $\mathfrak{S} \ominus \mathfrak{R}(X_0)$ . Thus  $f$  is in  $\mathfrak{R}(X_0)$  and  $f = X_0^{-1}f$ . Since the inverse of  $I - X_0$  exists, it follows that  $f = 0$ . Similarly, if  $X_0$  has range  $\mathfrak{S}_0$ , we have  $f = (X_0^{-1})^*f$  and a parallel argument leads again to the conclusion  $f = 0$ . Thus  $H \equiv i(I + X_0)(I - X_0)^{-1}$  is symmetric in  $\mathfrak{S}_0$  and, since either the domain or the range of  $X_0$  is identically  $\mathfrak{S}_0$ ,  $H$  is maximal as well. Furthermore, when  $\mathfrak{N}$  is hypermaximal  $W$ -symmetric,  $X_0$  has domain and range identically  $\mathfrak{S}_0$  so that in this case  $H$  is self-adjoint.

Next let us suppose that  $\mathfrak{S} = \mathfrak{S}_0 \oplus \mathfrak{D}_1(X)$  and that  $H$  is maximal symmetric in  $\mathfrak{S}_0$ . Then  $\mathfrak{N} = \mathfrak{B}(H) \oplus \mathfrak{B}$  where  $\mathfrak{B}$  is the manifold in  $\mathfrak{M} = \mathfrak{S} \oplus \mathfrak{S}$  whose elements are of the form  $\{0, h\}$ ,  $h$  in  $\mathfrak{D}_1(X)$ . Thus  $\mathfrak{M} \ominus \mathfrak{N} = W\mathfrak{B}(H^*) \oplus W\mathfrak{B}$ , and in view of Theorem 2.8 this equation reveals immediately that  $\mathfrak{N}$  is maximal  $W$ -symmetric. Furthermore, if  $H = H^*$ , then  $\mathfrak{B}(H^*) = \mathfrak{B}(H)$ , and  $\mathfrak{N}$  is clearly hypermaximal  $W$ -symmetric.

We bring this section to a close with a simple theorem which is revealed later as of fundamental importance.

**THEOREM 2.11.** *Let  $A$  be a reduction operator with domain in the graph of the adjoint  $H^*$  of a symmetric transformation  $H$  in a Hilbert space  $\mathfrak{S}$ . Let  $\mathfrak{M}$  be the range-space of  $A$ , and let  $W$  be the unitary transformation associated with  $A$  by Definition 1.1. Let  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  be the characteristic manifolds of  $W$  for the characteristic values  $+i$  and  $-i$ , respectively. Let  $U$  be the transformation in  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$  which takes  $\{f_1, f_2, h\}$  into  $\{f_2, -f_1, Wh\}$ . Then  $U$  is a unitary transformation in  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$  and  $U^2 + I = 0$ . The graph  $\mathfrak{B}(A)$  of  $A$ , consisting of all vectors of the form  $\{f, H^*f, Af\}$  in  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$  is a hypermaximal  $U$ -symmetric manifold. The characteristic manifolds of  $U$  for the characteristic values  $+i$  and  $-i$  are, respectively, the manifold  $\mathfrak{E}^+$  consisting of all elements of the form  $\{f, if, h_+\}$ , where  $f$  is in  $\mathfrak{S}$ ,  $h_+$  in  $\mathfrak{M}^+$ , and the manifold  $\mathfrak{E}^-$  of all elements of the form  $\{f, -if, h_-\}$ , where  $f$  is in  $\mathfrak{S}$  and  $h_-$  in  $\mathfrak{M}^-$ .*

Since it is an immediate consequence of Definition 1.1 that

$$U\mathfrak{B}(A) = (\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}) \ominus \mathfrak{B}(A),$$

it is necessary only to establish that  $U$  has the properties stated in the theorem. That  $U^2 + I \equiv O$  is readily verified:

$$U^2\{f_1, f_2, h\} = U\{f_2, -f_1, Wh\} = \{-f_1, -f_2, W^2h\}$$

and  $W^2h = -h$ . From inspection it is evident that  $\mathfrak{E}^+$  and  $\mathfrak{E}^-$  are at least subsets of the indicated characteristic manifolds of  $U$ ; that they must be identically those manifolds is an immediate consequence of the easily proved equality

$$\mathfrak{E}^+ + \mathfrak{E}^- = \mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}.$$

**4. Real  $W$ -symmetric manifolds.** Just as it is desirable to consider the special properties of symmetric transformations which are real with respect to a conjugation, so here it is desirable to consider "real"  $W$ -symmetric manifolds.

We recall that a conjugation  $J$  in a Hilbert space or unitary space  $\mathfrak{M}$  is a one-to-one transformation of  $\mathfrak{M}$  into itself such that

$$J^2 \equiv I, (Jf, Jg) = (\overline{f}, \overline{g}) = (g, f),$$

and that a transformation  $T$  in  $\mathfrak{M}$  is said to be real with respect to  $J$  if it permutes with  $J$ .†

**DEFINITION 2.4.** A linear manifold  $\mathfrak{N}$  in a unitary or Hilbert space  $\mathfrak{M}$  is said to be real with respect to a conjugation if it contains  $Jf$  whenever it contains  $f$ .

**THEOREM 2.12.** If a linear manifold  $\mathfrak{N}$  in  $\mathfrak{M}$  is real with respect to a conjugation  $J$ , then the closed linear manifold which  $\mathfrak{N}$  determines is real with respect to  $J$ .

For a linear manifold  $\mathfrak{N}$  to be real with respect to  $J$  it is necessary and sufficient that (1)  $J\mathfrak{N} = \mathfrak{N}$ ; if  $\mathfrak{N}$  is also closed, either of the following conditions is necessary and sufficient: (2)  $\mathfrak{M} \ominus \mathfrak{N}$  is real with respect to  $J$ , (3)  $(\mathfrak{M} \ominus \mathfrak{N}) = J(\mathfrak{M} \ominus \mathfrak{N})$ .

Since a conjugation  $J$  is evidently continuous, it follows at once that the closed linear manifold  $\overline{\mathfrak{N}}$  determined by a linear manifold  $\mathfrak{N}$  is real with respect to  $J$ , if  $\mathfrak{N}$  itself is.

If  $\mathfrak{N}$  is real with respect to  $J$ , then  $J\mathfrak{N}$  is a subset of  $\mathfrak{N}$ ; if it is a proper subset, then  $J^2\mathfrak{N}$  is also a proper subset of  $\mathfrak{N}$ . The latter is clearly impossible, however, since  $J^2 \equiv I$ . Thus the condition  $J\mathfrak{N} = \mathfrak{N}$  is necessary for the reality of  $\mathfrak{N}$  with respect to  $J$ ; that it is sufficient is obvious. For  $\mathfrak{N}$  closed, the neces-

† Stone, pp. 357-365.

sity and sufficiency of the condition (3) now follow from the fact that  $(Jf, Jg) = \overline{(f, g)}$  for all  $f$  in  $\mathfrak{N}$  and  $g$  in  $\mathfrak{M} \ominus \mathfrak{N}$ , and the fact that  $J$  takes  $\mathfrak{N}$  into itself and (1). Since (2) is, by (1), equivalent to (3), the proof is complete.

LEMMA 2.1. *A unitary transformation  $W$ ,  $W^2 + I = 0$ , permutes with a conjugation  $J$  if and only if  $\mathfrak{M}^- = J\mathfrak{M}^+$  and  $\mathfrak{M}^+ = J\mathfrak{M}^-$ .*

The proof is immediate.

THEOREM 2.13. *Let  $\mathfrak{N}$  be a closed  $W$ -symmetric manifold in  $\mathfrak{M}$ , and let  $V \equiv V(\mathfrak{N})$ . Then for  $\mathfrak{N}$  to be real with respect to a conjugation  $J$  which permutes with  $W$ , it is necessary and sufficient that the identity  $V \equiv JV^{-1}J$  hold.*

We observe first that  $J\mathfrak{N}$  is  $W$ -symmetric, since the relations

$$W\mathfrak{N} \subseteq \mathfrak{M} \ominus \mathfrak{N}, \quad J\mathfrak{N} \subseteq \mathfrak{M} \ominus J\mathfrak{N}, \quad WJ\mathfrak{N} \subseteq \mathfrak{M} \ominus J\mathfrak{N}$$

are equivalent. Also, if  $\mathfrak{N} = \mathfrak{N}(I - V)$ , then  $J\mathfrak{N} = \mathfrak{N}(I - JV^{-1}J)$ , and  $JV^{-1}J$  is isometric with domain in  $\mathfrak{M}^+$  and range in  $\mathfrak{M}^-$  by Lemma 2.1. Hence, by Theorem 2.2,  $\mathfrak{N} = J\mathfrak{N}$  if and only if  $V \equiv JV^{-1}J$ , and the proof is complete.

THEOREM 2.14. *Let  $\mathfrak{M} = \mathfrak{S} \oplus \mathfrak{S}$ , let  $W$  be as in Theorem 2.7, and let  $J_0$  be a conjugation in  $\mathfrak{S}$ . Then the transformation  $J$  in  $\mathfrak{M}$  defined by the equation  $J\{f, g\} = \{J_0f, J_0g\}$  is a conjugation which permutes with  $W$ . A closed linear  $W$ -symmetric manifold  $\mathfrak{N}$  in  $\mathfrak{M}$  is real with respect to  $J$  if and only if the manifold  $\mathfrak{D}_1(X)$  and the transformation  $H$  associated with  $\mathfrak{N}$  by Theorem 2.10 are real with respect to  $J_0$ .*

The first assertion is a direct consequence of Lemma 2.1.

If  $H$  is real with respect to  $J_0$ , then  $\mathfrak{B}(H)$  is real with respect to  $J$ , by definition. Furthermore, the manifold  $\mathfrak{D} + \mathfrak{D}_1(X)$  in  $\mathfrak{M} = \mathfrak{S} \oplus \mathfrak{S}$  is clearly real with respect to  $J$  if  $\mathfrak{D}_1(X)$  is real with respect to  $J_0$ . Thus, since

$$\mathfrak{N} = \mathfrak{B}(H) + (\mathfrak{D} + \mathfrak{D}_1(X)),$$

$\mathfrak{N}$  is real with respect to  $J$  when  $H$  and  $\mathfrak{D}_1(X)$  are real with respect to  $J_0$ .

On the other hand, suppose  $\mathfrak{N}$  is real with respect to  $J$ . Then

$$J\mathfrak{B}(H) + (\mathfrak{D} + J_0\mathfrak{D}_1(X)) = \mathfrak{B}(H) + (\mathfrak{D} + \mathfrak{D}_1(X)).$$

But  $J\mathfrak{B}(H) = \mathfrak{B}(J_0HJ_0)$  and  $J_0\mathfrak{D}_1(X)$  is orthogonal to the range of  $J_0HJ_0$ . Furthermore  $J_0HJ_0$  is readily shown to be a Hermitian transformation. But according to Theorem 2.10,  $H$  and  $\mathfrak{D}_1(X)$  orthogonal to the range of  $H$  are uniquely determined by  $\mathfrak{N}$ ; so we have

$$J_0HJ_0 \equiv H, \quad J_0\mathfrak{D}_1(X) = \mathfrak{D}_1(X).$$

Consequently  $H$  and  $\mathfrak{D}_1(X)$  are real with respect to  $J_0$ .

## CHAPTER III. REDUCTION OPERATORS

1. **An alternative definition.** In this chapter we examine in detail the situation described in Definition 1.1, with particular attention to the structure of transformations  $A$  satisfying the conditions of that definition.

Before proceeding, we introduce the following notations whose meanings remain fixed throughout the chapter:

**DEFINITION 3.1.**  $H$  is a closed symmetric transformation in Hilbert space  $\mathfrak{H}$ , with domain  $\mathfrak{D}$ ;  $\mathfrak{D}^*$  is the domain of  $H^*$ ;  $\mathfrak{B}$  and  $\mathfrak{B}^*$  are the graphs of  $H$  and  $H^*$ , respectively;  $\mathfrak{D}^+$  and  $\mathfrak{D}^-$  are the characteristic manifolds of  $H^*$  for the characteristic values  $+i$  and  $-i$ , respectively;  $\mathfrak{B}^+$  ( $\mathfrak{B}^-$ ) is the subset of  $\mathfrak{B}^*$  defined as follows:  $\{f, H^*f\}$  in  $\mathfrak{B}^*$  is in  $\mathfrak{B}^+$  ( $\mathfrak{B}^-$ ) if and only if  $f$  is in  $\mathfrak{D}^+$  ( $\mathfrak{D}^-$ ).

We recall from Theorem 2.8 that  $\mathfrak{B}^* \ominus \mathfrak{B} = \mathfrak{B}^+ + \mathfrak{B}^-$ .

**THEOREM 3.1.** If  $A$  is a reduction operator for  $H^*$ , the contraction  $A_1$  of  $A$  with domain  $[(\mathfrak{B}^+ + \mathfrak{B}^-) \cdot \mathfrak{D}(A)]$  is a closed linear transformation with domain dense in  $\mathfrak{B}^+ + \mathfrak{B}^-$  and range dense in  $\mathfrak{M}$ , the range-space of  $A$ . The transformations  $A_1^*$  and  $A_1^{-1}$  exist and

$$(3.1) \quad A_1^* \equiv QA_1^{-1}W,$$

where  $Q$  is the transformation in  $\mathfrak{B}^+ + \mathfrak{B}^-$  which takes  $\{f, g\}$  into  $\{g, -f\}$ ,  $f$  and  $g$  being elements of  $\mathfrak{H}$ .

Conversely, if  $A_1$  is any closed linear transformation with domain dense in  $\mathfrak{B}^+ + \mathfrak{B}^-$  and range in a unitary or Hilbert space  $\mathfrak{M}$ , and if  $A_1^{-1}$  exists and satisfies (3.1) for some unitary transformation  $W$  in  $\mathfrak{M}$ , then the closed linear transformation  $A$  with domain  $\mathfrak{D}(A_1) + \mathfrak{B}$  which is equal to 0 on  $\mathfrak{B}$  and to  $A_1$  on  $\mathfrak{D}(A_1)$ , is a reduction operator for  $H^*$ .

We note first that  $\mathfrak{D}(A_1)$  is dense in  $\mathfrak{B}^+ + \mathfrak{B}^-$ , since  $\mathfrak{B}$  is a closed linear manifold in  $\mathfrak{D}(A)$  and  $\mathfrak{D}(A)$  is dense in  $\mathfrak{B}^*$ . Thus the existence of  $A_1^*$  is assured. Furthermore, the existence of  $A_1^{-1}$  is assured by Theorem 1.1, and the relation  $A_1^* \supseteq QA_1^{-1}W$  follows at once from Definition 1.1.

Hence we have only to show that  $QA_1^{-1}W \supseteq A_1^*$ ; to do this we take direct recourse to the definition of the adjoint. Accordingly, we consider an arbitrary vector  $\{f, H^*f, h\}$  in  $(\mathfrak{B}^+ + \mathfrak{B}^-) \oplus \mathfrak{M}$  such that the equation

$$(g, f) + (H^*g, H^*f) - (Ag, H^*g, h) = 0$$

is satisfied for all  $\{g, H^*g\}$  in  $\mathfrak{D}(A_1)$ . Since  $\{f, H^*f\}$  is by hypothesis in  $\mathfrak{B}^* \ominus \mathfrak{B}$ , the equation is satisfied for all  $\{g, H^*g\}$  in  $\mathfrak{D}(A_1)$  if and only if it is satisfied for all  $\{g, H^*g\}$  in  $\mathfrak{D}(A)$ . Thus, by Definition 1.1, we must have  $\{f, H^*f\} = \{H^*k, -k\}$ ,  $-h = WA_1\{k, H^*k\}$ . Consequently, since  $W \equiv -W^{-1}$ ,

$$\{k, H^*k\} = -A_1^{-1}W^{-1}h = A_1^{-1}Wh.$$

But  $\{f, H^*f\} = A_1^*h$  and  $\{f, H^*f\} = Q\{k, H^*k\}$ . Therefore, whenever  $h$  is in the domain of  $A_1^*$ ,  $Wh$  is in the domain of  $A_1^{-1}$  and  $A_1^*h = QA_1^{-1}Wh$ . Thus  $A_1^* \subseteq QA_1^{-1}W$ .

If, on the other hand,  $A_1$  is given with the properties specified in the theorem, it is readily shown that its linear extension  $A$  defined by the relations  $A = A_1$  on  $\mathfrak{D}(A_1)$ ,  $A = 0$  on  $\mathfrak{B}$ , satisfies the conditions of Definition 1.1 and so is a reduction operator for  $H^*$ .

Throughout the remainder of this paper, the subscript 1 attached to the symbol for a reduction operator indicates the contraction defined in Theorem 3.1. With this convention, we may take as the definition of a reduction operator the identity (3.1). The latter has the advantage of greater transparency; in particular, it reveals at once the following important information:

**THEOREM 3.2.** *The range of a reduction operator  $A$  for  $H^*$  is the entire range-space  $\mathfrak{M}$  of  $A$  if and only if  $\mathfrak{M}$  is a unitary space or  $A$  is bounded.*

If  $\mathfrak{M}$  is unitary,  $A$  is necessarily bounded, so we may suppress consideration of the dimension number of  $\mathfrak{M}$ .

Obviously  $A$  is bounded if and only if  $A_1$  is and thus if and only if  $A_1^*$  is. But  $A_1^*$  being closed is bounded if and only if its domain is  $\mathfrak{M}$ , and, by the identity (3.1), its domain is  $\mathfrak{M}$  if and only if the domain of  $A_1^{-1}$  (that is, the range of  $A$ ) is  $\mathfrak{M}$ .

On the basis of Theorem 3.2, it would be possible to proceed at once to the solution of the problem proposed at the end of Chapter I for the case that  $A$  is bounded. For unbounded reduction operators, however, more elaborate analysis is necessary.

**2. Characterization of all reduction operators.** We give now a characterization of all reduction operators which, except for minor points of detail, is due to M. H. Stone.

**THEOREM 3.3.** *If  $A$  is a reduction operator for  $H^*$ , with range-space  $\mathfrak{M}$ , and  $X$  is an arbitrary isometric transformation with domain  $\mathfrak{M}$  and range  $\mathfrak{N}$ , then  $C \equiv XA$  is a reduction operator for  $H^*$ , with range-space  $\mathfrak{N}$ . The unitary transformation in  $\mathfrak{N}$  associated with  $C$  by Definition 1.1 is  $U \equiv XWX^{-1}$ . The characteristic manifolds of  $U$  for the characteristic values  $+i$  and  $-i$  are  $X\mathfrak{M}^+$  and  $X\mathfrak{M}^-$ , respectively, where  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  are the characteristic manifolds of  $W$  for the characteristic values  $+i$  and  $-i$ , respectively.*

We set  $C_1 \equiv XA_1$ , obtaining  $A_1 \equiv X^{-1}C_1$ ,  $A_1^* \equiv C_1^*X$ ,  $A_1^{-1} \equiv C_1^{-1}X$ . Thus from the identity (3.1) we have  $C_1^*X \equiv QC_1^{-1}XW$ , or  $C_1^* \equiv QC_1^{-1}XWX^{-1}$ , which by Theorem 3.1 implies that  $C \equiv XA$  is a reduction operator for  $H^*$ .

To prove the final assertions of the theorem, we make use of the identities  $UX = XW$ ,  $X^{-1}U = WX^{-1}$ . If  $Wh = ah$ , we have, by the first identity,  $UXh = Xah = aXh$ . On the other hand, if  $Uh = ah$ , we have, by the second identity,  $aX^{-1}h = X^{-1}ah = WX^{-1}h$ . Setting  $a$  equal first to  $i$  and then to  $-i$ , we obtain at once the desired result.

**DEFINITION 3.2.** If  $A$  and  $C$  are reduction operators for  $H^*$ , with range-spaces  $\mathfrak{M}$  and  $\mathfrak{N}$ , respectively,  $C$  is said to be equivalent to  $A$ ,  $C \sim A$ , if there exists an isometric transformation  $X$  with domain  $\mathfrak{M}$  and range  $\mathfrak{N}$  such that  $C \equiv XA$ .

**THEOREM 3.4.** Let  $A$ ,  $C$ , and  $D$  be reduction operators for  $H^*$ . Then  $C \sim A$  implies  $A \sim C$ , and  $D \sim C$ ,  $C \sim A$  imply  $D \sim A$ .

The proof is immediate.

**DEFINITION 3.3.** The class of all reduction operators for  $H^*$  which are equivalent to a single reduction operator  $A$  is called the equivalence class of  $A$ , or merely an equivalence class.

We can now characterize all reduction operators for  $H^*$  by characterizing a representative member of each equivalence class. Our method is to show that  $(C_1^*C_1)^{1/2} \equiv (A_1^*A_1)^{1/2}$  whenever  $A$  and  $C$  are equivalent and that  $B_1 \equiv (A_1^*A_1)^{1/2}$  defines, in accordance with the second paragraph of Theorem 3.1, a reduction operator  $B$  for  $H^*$ , which is equivalent to  $A$ . Thus  $B$ , or  $B_1$ , completely determines the equivalence class of  $A$ . Our characterization of  $B_1$  is based on the following theorem:

**THEOREM 3.5.** Let  $\mathfrak{Q}$  be a Hilbert or unitary space. Let  $T$  be a nonnegative definite self-adjoint transformation in  $\mathfrak{Q}$  such that  $T^{-1}$  exists and satisfies the identity  $T \equiv QT^{-1}R$ , where  $Q$  and  $R$  are unitary transformations in  $\mathfrak{Q}$  such that  $Q^2 \equiv R^2 \equiv -I$ . Then  $R \equiv Q^{-1}$ , and the resolution of the identity  $E(\lambda)$  in  $\mathfrak{Q}$  associated with  $T$  has the following properties:

- (1)  $E(\lambda) \equiv O$  for  $\lambda \leq 0$ ;
- (2)  $E(\lambda) \equiv Q(I - E(1/\lambda - 0))Q^{-1}$  for  $0 < \lambda < \infty$ ;
- (3) the range  $\mathfrak{N}$  of  $E(1 - 0)$  is a  $Q$ -symmetric manifold;
- (4) the range of  $I - E(1)$  is  $Q\mathfrak{N}$ ;
- (5) the range  $\mathfrak{P}$  of  $E(1) - E(1 - 0)$  is of the form  $\mathfrak{P} = \mathfrak{N}_1 + \mathfrak{N}_2$ , where  $\mathfrak{N}_1$  and  $\mathfrak{N}_2$  are closed linear manifolds belonging to the characteristic manifolds of  $Q$  for the characteristic values  $i$  and  $-i$ , respectively.

Let  $\mathfrak{N}$  be an arbitrary closed  $Q$ -symmetric manifold in  $\mathfrak{Q}$ , and let  $E_{\mathfrak{N}}(\lambda)$  be a resolution of the identity in  $\mathfrak{N}$  such that  $E_{\mathfrak{N}}(\lambda) \equiv O$ ,  $\lambda \leq 0$ ;  $E_{\mathfrak{N}}(1 - 0) \equiv I$ . Then the conditions

$$(6) \quad E(\lambda) = E_{\mathfrak{N}}(\lambda)E_{\mathfrak{N}}, \quad \text{for } \lambda < 1;$$

$$(7) \quad E(1) = E_{\mathfrak{N}},$$

where  $\mathfrak{U} = \mathfrak{E} \ominus Q\mathfrak{N}$ ;

$$(8) \quad E(\lambda) = Q(I - E(1/\lambda - 0))Q^{-1}, \quad \text{for } 1 < \lambda < \infty,$$

define a resolution of the identity  $E(\lambda)$  in  $\mathfrak{E}$  whose associated self-adjoint transformation  $T$  is nonnegative definite and has an inverse such that  $T \equiv QT^{-1}Q^{-1}$ .

Since  $T \equiv QT^{-1}R$  is self-adjoint, we have also  $T \equiv R^{-1}T^{-1}Q^{-1}$ , the identity  $(QT^{-1}R)^* \equiv R^{-1}T^{-1}Q^{-1}$  being easily established. Hence  $T^2 \equiv QT^{-2}Q^{-1}$ , or  $T^2 \equiv (QT^{-1}Q^{-1})^2$ . As  $QT^{-1}Q^{-1}$  and  $T$  are both nonnegative definite, we conclude that they are equal and thus that  $R = Q^{-1}$  on  $\mathfrak{D}(T)$ . Since the latter manifold is dense in  $\mathfrak{E}$ , and  $R$  and  $Q^{-1}$  are both bounded, we have  $R = Q^{-1}$ .

Now let  $E(\lambda)$  be the resolution of the identity associated with  $T$ . Then the fact that  $T$  is nonnegative definite and has an inverse has as a consequence (1) of the theorem. To prove (2) we first recall that the unitary equivalence of two self-adjoint transformations is a necessary and sufficient condition for the unitary equivalence of the corresponding resolutions of the identity for each value of the real parameter  $\lambda$  which appears in the definition of both.† Since the resolution of the identity  $F(\lambda)$  for  $T^{-1}$  is obtained from that for  $T$  according to the equations

$$\begin{aligned} F(\lambda) &\equiv 0, & \lambda &\leq 0, \\ F(\lambda) &\equiv I - E(1/\lambda - 0), & 0 < \lambda < \infty, \end{aligned}$$

the unitary equivalence of  $T$  and  $T^{-1}$  under the unitary transformation  $Q$  yields the relation (2). From (2) we obtain at once  $E(1-0) \equiv Q(I - E(1))Q^{-1}$  which implies (4); and, since the range of  $I - E(1)$  is orthogonal to the range of  $E(1-0)$ , (3) now follows. In order to establish (5) we have only to observe that  $E(1) - E(1-0)$  has range  $\mathfrak{E} \ominus (\mathfrak{N} + Q\mathfrak{N})$  and then apply Theorem 2.2.

To prove the second part of the theorem we note first that the self-adjoint transformation  $T$  associated with the resolution of the identity defined by the relations (6), (7), and (8) has an inverse whose resolution of the identity  $F(\lambda)$  is obtained from  $E(\lambda)$  in the manner indicated above. Thus we have  $E(\lambda) \equiv QF(\lambda)Q^{-1}$  which, according to a theorem already indicated, implies  $T \equiv QT^{-1}Q^{-1}$ .

**THEOREM 3.6.** *Each equivalence class  $\mathcal{A}$  of reduction operators for  $H^*$  contains one and only one operator  $B$  with range-space  $\mathfrak{B}^+ + \mathfrak{B}^-$  such that  $B_1$  is a nonnegative definite self-adjoint transformation in  $\mathfrak{B}^+ + \mathfrak{B}^-$ . If  $B$  is such a reduction operator,  $B_1 \equiv QB_1^{-1}Q^{-1}$ . If  $A$  is an arbitrary reduction operator in the equivalence class of  $B$ , then  $A^*A \equiv B^2$  and  $A_1^*A_1 \equiv B_1^2$ .*

† Stone, Theorem 7.1.

Let the class  $\mathcal{A}$  be given, and let  $C$  belong to  $\mathcal{A}$ . Then the transformation  $C_1^*C_1$  is a nonnegative definite self-adjoint transformation and has a unique nonnegative definite square root  $B_1$ ,  $B_1^2 \equiv C_1^*C_1$ , such that  $B_1 \equiv XC_1$ , where  $X$  is an isometric transformation with domain the closure of the range of  $C_1$  and range the closure of the range of  $C_1^*$ .<sup>†</sup> Thus  $X$  is an isometric transformation with domain the range-space  $\mathfrak{M}$  of  $C$  and range  $\mathfrak{B}^+ + \mathfrak{B}^-$ . By Theorem 3.3,  $B \equiv XC$  belongs to  $\mathcal{A}$ . Since  $B_1 \equiv QB_1^{-1}XWX^{-1}$ , where  $W$  is the unitary operator in  $\mathfrak{M}$  associated with  $C$ , we have by Theorem 3.5,  $B_1 \equiv QB_1^{-1}Q^{-1}$ . Now let  $A$  be any other member of the class  $\mathcal{A}$ . Then  $A_1 \equiv YC_1$  and  $A_1^* \equiv C_1^*Y^{-1}$ , where  $Y$  is isometric. Thus  $A_1^*A_1 \equiv B_1^2$ , from which the equation  $A^*A \equiv B^2$  follows at once, since  $A_1^* \equiv A^*$ . In particular, if  $A_1$  is self-adjoint in  $\mathfrak{B}^+ + \mathfrak{B}^-$ , we have  $A_1^2 \equiv B_1^2$ . Consequently, if  $A_1$  is also nonnegative definite, we must have  $A_1 \equiv B_1$ . Thus  $B$  is unique, and the proof is complete.

**DEFINITION 3.4.** If  $A$  is a reduction operator for  $H^*$ ,  $E_A(\lambda)$  denotes the resolution of the identity in  $\mathfrak{B}^+ + \mathfrak{B}^-$  associated with  $(A_1^*A_1)^{1/2}$ .

Theorems 3.3, 3.5, and 3.6 evidently provide a constructive characterization of all reduction operators for  $H^*$ . This characterization, since it allows us to study an arbitrary reduction operator  $A$  in terms of the self-adjoint transformation  $B_1 \equiv (A_1^*A_1)^{1/2}$  whose properties are described in Theorem 3.5, leads to comparatively simple proofs of many theorems which are otherwise established only with considerable difficulty. In particular, it reveals the effect of the deficiency index of  $H$  on the reduction operators for  $H^*$ .

**THEOREM 3.7.** Let  $A$  be a reduction operator for  $H^*$ , and let  $\mathfrak{M}$  and  $W$  have the same meanings as in Definition 1.1,  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  the same meanings as in Theorem 2.2. Let  $m$  and  $n$  be the dimension numbers of  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$ , respectively. Then  $(n, m)$  is the deficiency index of  $H$ .

A necessary and sufficient condition that there be unbounded reduction operators for  $H^*$  is that  $H$  have the deficiency index  $(\aleph_0, \aleph_0)$ .

Evidently  $\mathfrak{B}^-$  and  $\mathfrak{B}^+$  are the characteristic manifolds of  $Q^{-1}$  for the characteristic values  $i$  and  $-i$ , respectively. But then, by Theorems 3.6 and 3.3, we have  $\mathfrak{B}^- = X\mathfrak{M}^+$  and  $\mathfrak{B}^+ = X\mathfrak{M}^-$ , where  $X$  is an isometric transformation with domain  $\mathfrak{M}$  and range  $\mathfrak{B}^+ + \mathfrak{B}^-$ . Thus  $\mathfrak{B}^-$  has the dimension number  $m$ ,  $\mathfrak{B}^+$  the dimension number  $n$ . Since  $\mathfrak{B}^-$  and  $\mathfrak{B}^+$  clearly have the same dimension numbers as  $\mathfrak{D}^-$  and  $\mathfrak{D}^+$ , respectively, the deficiency index of  $H$  is  $(n, m)$ .

To prove the second assertion of the theorem, we observe first that a reduction operator  $A$  is unbounded if and only if  $B_1 \equiv (A_1^*A_1)^{1/2}$  is unbounded. Now let  $\mathfrak{R}$  be the range of  $E_A(1-0)$ . Then, by Theorem 3.5, the range of

<sup>†</sup> Murray, Theorem 1.24.

$I - E_A(1)$  is  $Q\mathfrak{N}$  and  $\mathfrak{N}$  is  $Q$ -symmetric. Furthermore, on  $(\mathfrak{B}^+ + \mathfrak{B}^-) \ominus (\mathfrak{N} + Q\mathfrak{N})$ ,  $B_1 = I$ . Thus  $B_1$  induces a bounded transformation in  $(\mathfrak{B}^+ + \mathfrak{B}^-) \ominus Q\mathfrak{N}$ , and so is unbounded if and only if it induces an unbounded transformation in  $Q\mathfrak{N}$ . Since it is evident that  $B_1$  can be constructed with this property if and only if  $\mathfrak{N}$  has the dimension number  $\aleph_0$ , and since, by Theorem 2.2,  $\mathfrak{N}$  can be chosen with that dimension number if and only if both of the characteristic manifolds of  $Q$  are Hilbert spaces, the proof is complete.

**3. The graph of a reduction operator.** For further study of reduction operators, we now call into play Theorem 2.11. We preserve the meanings of all the symbols introduced in that theorem.

**THEOREM 3.8.** *The space  $\mathfrak{B} + \mathfrak{D}$  in  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$  is  $U$ -symmetric and  $\mathfrak{B} + \mathfrak{D} \subseteq \mathfrak{B}(A)$ , the equality holding if and only if  $H \equiv H^*$ . Let  $Z$  be the isometric transformation with domain  $\mathfrak{E}^+$  and range  $\mathfrak{E}^-$  associated with  $\mathfrak{B} + \mathfrak{D}$  by Theorem 2.2, and let  $Y$  be the unitary transformation of  $\mathfrak{E}^+$  into  $\mathfrak{E}^-$  associated with  $\mathfrak{B}(A)$  in accordance with Theorems 2.11 and 2.2. Then  $Z \subseteq Y$ , the equality holding if and only if  $H \equiv H^*$ . The manifolds  $\mathfrak{E}^+$  and  $\mathfrak{E}^-$  admit the decompositions*

$$(1) \quad \mathfrak{E}^+ = \mathfrak{D}(Z) \oplus \mathfrak{B}^+ \oplus \mathfrak{M}^+,$$

$$(2) \quad \mathfrak{E}^- = \mathfrak{R}(Z) \oplus \mathfrak{B}^- \oplus \mathfrak{M}^-,$$

where the three component manifolds in each case are mutually orthogonal, and  $\mathfrak{B} \oplus \mathfrak{M}^- = Y(\mathfrak{B}^+ \oplus \mathfrak{M}^+)$ .

That the relation  $\mathfrak{B} + \mathfrak{D} \subseteq \mathfrak{B}(A)$  holds is obvious, since  $\mathfrak{B}$  is the manifold of zeros of  $A$ . Hence, by Theorem 2.2,  $Z \subseteq Y$ , and  $\mathfrak{B} + \mathfrak{D} = \mathfrak{D}(A)$  if and only if  $Z = Y$ . But if  $\mathfrak{B} + \mathfrak{D} = \mathfrak{D}(A)$ , then  $A \equiv O$ , and conversely. Since  $A \equiv O$  if and only if  $H \equiv H^*$ , the first two assertions of the theorem are established.

As we observed in Theorem 2.11, the manifold  $\mathfrak{E}^+$  consists of all vectors in  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$  of the form  $\{f, if, h_+\}$ , where  $h_+$  is in  $\mathfrak{M}^+$ , and  $\mathfrak{E}^-$  of all vectors of the form  $\{f, -if, h_-\}$ , where  $h_-$  is in  $\mathfrak{M}^-$ . Since, if

$$X \equiv (H - iI)(H + iI)^{-1},$$

we have  $\mathfrak{D}(X) = \mathfrak{S} \ominus \mathfrak{D}^+$  and  $\mathfrak{R}(X) = \mathfrak{S} \ominus \mathfrak{D}^-$ , it is readily shown, by reason of the relation between  $X$  and  $Z$  made evident by Theorem 2.7, that  $\mathfrak{B}^+ \oplus \mathfrak{M}^+ = \mathfrak{E}^+ \ominus \mathfrak{D}(Z)$  and that  $\mathfrak{B}^- \oplus \mathfrak{M}^- = \mathfrak{E}^- \ominus \mathfrak{R}(Z)$ . From these equations the assertions of the theorem concerning the decomposition of  $\mathfrak{E}^+$  and  $\mathfrak{E}^-$  follow.

Finally, the equation  $\mathfrak{B} \oplus \mathfrak{M}^- = Y(\mathfrak{B}^+ \oplus \mathfrak{M}^+)$  is a consequence of the unitary character of  $Y$  and the facts already established.

**THEOREM 3.9.** *Let  $\mathfrak{B}^+(A)$ ,  $\mathfrak{B}^-(A)$ ,  $\mathfrak{B}_+(A)$ , and  $\mathfrak{B}_-(A)$  be the subsets of  $\mathfrak{B}(A_1)$  defined as follows:  $\{f, H^*f, Af\}$  in  $\mathfrak{B}(A_1)$  is in  $\mathfrak{B}^+(A)$  if  $f$  is in  $\mathfrak{D}^+$ , and*

in  $\mathfrak{B}^-(A)$  if  $f$  is in  $\mathfrak{D}^-$ ;  $\{f, H^*f, Af\}$  is in  $\mathfrak{B}_+(A)$  if  $Af$  is in  $\mathfrak{M}^+$ , and in  $\mathfrak{B}_-(A)$  if  $Af$  is in  $\mathfrak{M}^-$ . Then

$$(1) \quad \mathfrak{B}^+(A) = (I - Y^{-1})\mathfrak{M}^-,$$

$$(2) \quad \mathfrak{B}^-(A) = (I - Y)\mathfrak{M}^+,$$

$$(3) \quad \mathfrak{B}_+(A) = (I - Y^{-1})\mathfrak{B}^-,$$

$$(4) \quad \mathfrak{B}_-(A) = (I - Y)\mathfrak{B}^+;$$

and

$$(5) \quad \mathfrak{B}(A_1) \ominus \mathfrak{B}^+(A) = \mathfrak{B}_+(A),$$

$$(6) \quad \mathfrak{B}(A_1) \ominus \mathfrak{B}^-(A) = \mathfrak{B}_-(A).$$

Thus

$$(7) \quad \mathfrak{B}(A) = \mathfrak{B} + \mathfrak{B}^+(A) + \mathfrak{B}_+(A) = \mathfrak{B} + \mathfrak{B}^-(A) + \mathfrak{B}_-(A).$$

From Theorem 3.8 it is evident that every element of  $\mathfrak{B}(A_1)$  can be written in the form  $(I - Y)\{f^+, if^+, h_+\}$  where  $f^+$  is in  $\mathfrak{D}^+$  and  $h_+$  is in  $\mathfrak{M}^+$ , or equivalently in the form  $(I - Y^{-1})\{f^-, -if^-, h_-\}$  where  $f^-$  is in  $\mathfrak{D}^-$ ,  $h_-$  in  $\mathfrak{M}^-$ . However, since  $Y^{-1}$  takes  $\mathfrak{B}^- + \mathfrak{M}^-$  into  $\mathfrak{B}^+ + \mathfrak{M}^+$ , it is clear that the projection of  $(I - Y^{-1})\{f^-, -if^-, h_-\}$  on  $\mathfrak{S} \oplus \mathfrak{S}$  is in  $\mathfrak{B}^+$ ; that is, that  $(I - Y^{-1})\{f^-, -if^-, h_-\}$  is in  $\mathfrak{B}^+(A)$  if and only if  $f^- = 0$ . Thus (1) is valid. Equations (2), (3), and (4) can be verified by entirely similar arguments. Equation (5) then follows from equations (1) and (3) and equation (2) of Theorem 3.8, while (6) follows from (2) and (4) and equation (1) of Theorem 3.8. The relations (7) are consequences of (5) and (6) and the fact that  $\mathfrak{B}(A_1) = \mathfrak{B}(A) \ominus (\mathfrak{B} + \mathfrak{D})$ .

It is worth while to point out here a second characterization of reduction operators which is suggested by Theorems 2.11, 3.8, and 3.9. The facts can be stated as follows: Let  $\mathfrak{S}$  be a Hilbert space,  $H$  a closed linear symmetric transformation in  $\mathfrak{S}$ . Let  $Q$  be the transformation in  $\mathfrak{S} \oplus \mathfrak{S}$  which takes  $\{f, g\}$  into  $\{g, -f\}$ , and let  $Z$  be the isometric transformation corresponding to the  $Q$ -symmetric manifold  $\mathfrak{B} = \mathfrak{B}(H)$  in accordance with Theorem 2.2. Let  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  be complex euclidean spaces with the same dimension numbers as  $\mathfrak{B}^-$  and  $\mathfrak{B}^+$ , respectively. Let  $Y$  be an arbitrary isometric transformation with domain  $\mathfrak{D}(Z) + \mathfrak{B}^+ + \mathfrak{M}^+$  and range  $\mathfrak{R}(Z) + \mathfrak{B}^- + \mathfrak{M}^-$  such that  $Y \supseteq Z$ , such that  $E_{\mathfrak{B}^-} Y h^+ = 0$  for  $h^+$  in  $\mathfrak{M}^+$  implies  $h^+ = 0$ , and such that

$$E_{\mathfrak{M}^-} Y \{f^+, if^+\} = 0$$

for  $\{f^+, if^+\}$  in  $\mathfrak{B}^+$  implies  $f^+ = 0$ . Then  $\mathfrak{R}(I - Y)$  is the graph of a reduction operator  $A$  for  $H^*$ . We leave the proof of these assertions to the reader.

**THEOREM 3.10.** *Let  $\mathfrak{B}_A^+ = \mathfrak{B}^+ \cdot \mathfrak{D}(A)$ ,  $\mathfrak{B}_A^- = \mathfrak{B}^- \cdot \mathfrak{D}(A)$ ,  $\mathfrak{M}_A^+ = \mathfrak{M}^+ \cdot \mathfrak{R}(A)$ ,  $\mathfrak{M}_A^- = \mathfrak{M}^- \cdot \mathfrak{R}(A)$ . Then  $\mathfrak{B}_A^+$ ,  $\mathfrak{B}_A^-$ ,  $\mathfrak{M}_A^+$ , and  $\mathfrak{M}_A^-$  are linear manifolds everywhere dense in  $\mathfrak{B}^+$ ,  $\mathfrak{B}^-$ ,  $\mathfrak{M}^+$ , and  $\mathfrak{M}^-$ , respectively. Each of the equations  $\mathfrak{B}_A^+ = \mathfrak{B}^+$ ,  $\mathfrak{B}_A^- = \mathfrak{B}^-$ ,  $\mathfrak{M}_A^+ = \mathfrak{M}^+$ ,  $\mathfrak{M}_A^- = \mathfrak{M}^-$ , is valid if and only if  $A$  is bounded. If  $V$  is the isometric transformation with domain  $\mathfrak{M}$  and range  $\mathfrak{B}^+ + \mathfrak{B}^-$  such that  $(A_1^* A_1)^{1/2} \equiv V A_1$ , then  $\mathfrak{B}_A^- = V \mathfrak{M}_A^+$ ,  $\mathfrak{B}_A^+ = V \mathfrak{M}_A^-$ .*

*The manifolds  $\mathfrak{B}_A^+$ ,  $\mathfrak{B}_A^-$ ,  $\mathfrak{M}_A^+$ , and  $\mathfrak{M}_A^-$  are characterized by the equations*

- (1)  $\mathfrak{B}_A^+ = E_{\mathfrak{S} \oplus \mathfrak{S}} \mathfrak{B}^+(A) = E_{\mathfrak{S} \oplus \mathfrak{S}} Y^{-1} \mathfrak{M}^-$ ,
- (2)  $\mathfrak{B}_A^- = E_{\mathfrak{S} \oplus \mathfrak{S}} \mathfrak{B}^-(A) = E_{\mathfrak{S} \oplus \mathfrak{S}} Y \mathfrak{M}^+$ ,
- (3)  $\mathfrak{M}_A^+ = E_{\mathfrak{M}} \mathfrak{B}_+(A) = E_{\mathfrak{M}} Y^{-1} \mathfrak{B}^-$ ,
- (4)  $\mathfrak{M}_A^- = E_{\mathfrak{M}} \mathfrak{B}_-(A) = E_{\mathfrak{M}} Y \mathfrak{B}^+$ ,

where each of the projections  $E_{\mathfrak{S} \oplus \mathfrak{S}}$  and  $E_{\mathfrak{M}}$  has domain  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{M}$ , and  $Y$  has the meaning assigned to it in Theorem 3.8.

Again we denote by  $\mathfrak{N}$  the range of  $E_A(1-0)$ , and by  $\mathfrak{P}$  the range of  $E_A(1) - E_A(1-0)$ . Then  $\mathfrak{N}$  and  $\mathfrak{P}$  both reduce  $B_1 \equiv (A_1^* A_1)^{1/2}$ , and in  $\mathfrak{P}$ ,  $B_1 = I$ , while in  $\mathfrak{N}$ ,  $B_1$  induces a bounded self-adjoint transformation. Furthermore, by Theorem 3.5,  $\mathfrak{P} = \mathfrak{B}_1^+ + \mathfrak{B}_1^-$ , where  $\mathfrak{B}_1^+$  and  $\mathfrak{B}_1^-$  are closed linear manifolds in  $\mathfrak{B}^+$  and  $\mathfrak{B}^-$ , respectively; in addition, applying the same theorem and Theorem 2.2, we have  $\mathfrak{P} = (\mathfrak{B}^+ + \mathfrak{B}^-) \ominus (\mathfrak{N} + Q\mathfrak{N})$  and  $\mathfrak{N} + Q\mathfrak{N} = \mathfrak{D}(X) + \mathfrak{R}(X)$ , where  $X$  is an isometric transformation and  $\mathfrak{D}(X) = \mathfrak{B}^+ \ominus \mathfrak{B}_1^+$ ,  $\mathfrak{R}(X) = \mathfrak{B}^- \ominus \mathfrak{B}_1^-$ ,  $\mathfrak{N}(I-X) = \mathfrak{N}$ .

Now let  $\mathfrak{N}_1$  be the range of the transformation  $B_2$  induced in  $\mathfrak{N}$  by  $B_1$ . Then, since  $B_2$  is self-adjoint and has an inverse,  $\mathfrak{N}_1$  is a linear manifold dense in  $\mathfrak{N}$ , and  $\mathfrak{D}(A_1) = \mathfrak{D}(B_1) = \mathfrak{P} + \mathfrak{N} + Q\mathfrak{N}_1$ . Hence

$$\begin{aligned}\mathfrak{D}(A) \cdot \mathfrak{B}^+ &= \mathfrak{D}(A_1) \cdot \mathfrak{B}^+ = \mathfrak{B}_1^+ + (\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^+, \\ \mathfrak{D}(A) \cdot \mathfrak{B}^- &= \mathfrak{D}(A_1) \cdot \mathfrak{B}^- = \mathfrak{B}_1^- + (\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^-.\end{aligned}$$

Consequently,  $\mathfrak{B}_A^+$  ( $\mathfrak{B}_A^-$ ) is dense in  $\mathfrak{B}^+$  ( $\mathfrak{B}^-$ ) if and only if  $(\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^+$  ( $(\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^-$ ) is dense in  $\mathfrak{D}(X)$  ( $\mathfrak{R}(X)$ ).

To prove that the latter condition is satisfied, we denote by  $X_1$  the isometric transformation from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that  $\mathfrak{N}_1 = \mathfrak{N}(I - X_1)$ . Then evidently  $Q\mathfrak{N}_1 = \mathfrak{N}(I + X_1)$ . Thus, since  $\mathfrak{N} \supseteq \mathfrak{N}(I - X_1)$ , we have

$$(\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^+ = \mathfrak{D}(X_1), \quad (\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^- = \mathfrak{R}(X_1),$$

and

$$\mathfrak{B}_A^+ = \mathfrak{B}_1^+ + \mathfrak{D}(X_1), \quad \mathfrak{B}_A^- = \mathfrak{B}_1^- + \mathfrak{R}(X_1).$$

But, since the closure of  $\mathfrak{N}_1$  is  $\mathfrak{N}$ , it follows from Theorem 2.2 that  $\bar{X}_1 \equiv X$ ,

and therefore that the closures of the domain and range of  $X_1$  are, respectively, the domain and range of  $X$ . Furthermore,  $\mathfrak{N}_1 = \mathfrak{N}$  if and only if  $B_2$  has a bounded inverse; that is, if and only if  $A$  is bounded. Thus  $X_1 \equiv X$  if and only if  $A$  is bounded. Therefore  $\mathfrak{B}_A^+$  and  $\mathfrak{B}_A^-$  are linear manifolds everywhere dense in  $\mathfrak{B}^+$  and  $\mathfrak{B}^-$ , respectively, and  $\mathfrak{B}_A^+ = \mathfrak{B}^+$ ,  $\mathfrak{B}_A^- = \mathfrak{B}^-$  if and only if  $A$  is bounded.

We now observe that  $\mathfrak{N}(B_1) = V\mathfrak{N}(A_1)$  and, since  $(\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^+ = (\mathfrak{N}_1 + Q\mathfrak{N}) \cdot \mathfrak{B}^+$  and  $(\mathfrak{N} + Q\mathfrak{N}_1) \cdot \mathfrak{B}^- = (\mathfrak{N}_1 + Q\mathfrak{N}) \cdot \mathfrak{B}^-$ , that

$$\mathfrak{D}(B_1) \cdot \mathfrak{B}^+ = \mathfrak{N}(B_1) \cdot \mathfrak{B}^-, \quad \mathfrak{D}(B_1) \cdot \mathfrak{B}^- = \mathfrak{N}(B_1) \cdot \mathfrak{B}^+.$$

Consequently, taking account of the relations  $\mathfrak{B}^- = V\mathfrak{M}^+$ ,  $\mathfrak{B}^+ = V\mathfrak{M}^-$ , we have  $\mathfrak{B}_A^+ = V\mathfrak{M}_A^-$ ,  $\mathfrak{B}_A^- = V\mathfrak{M}_A^+$ . Thus, since  $V$  is isometric, it follows that  $\mathfrak{M}_A^+$  and  $\mathfrak{M}_A^-$  are dense in  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$ , respectively, and that  $\mathfrak{M}_A^+ = \mathfrak{M}^+$ ,  $\mathfrak{M}_A^- = \mathfrak{M}^-$  if and only if  $A$  is bounded. Thus the facts stated in the first paragraph are all established.

The equations with which the theorem concludes are consequences of Theorems 2.11, 3.8, and 3.9, as the reader can readily verify.

**THEOREM 3.11.** *Let  $A$  be a reduction operator for  $H^*$ , and let  $Y$  have the same meaning as in Theorem 3.8. Let  $F$  be the transformation with domain  $\mathfrak{M}^+$  which is equal on its domain to  $E_{\mathfrak{M}}Y$ , and let  $G$  be the transformation with domain  $\mathfrak{B}^+$  which is equal on its domain to  $E_{\mathfrak{B} \oplus \mathfrak{B}}Y$ . Then  $F$  and  $G$  are bounded linear transformations with ranges in  $\mathfrak{M}^-$  and  $\mathfrak{B}^-$ , respectively. The adjoints  $F^*$  and  $G^*$ , with the respective domains  $\mathfrak{M}^-$  and  $\mathfrak{B}^-$ , are equal on their domains to  $E_{\mathfrak{M}}Y^{-1}$  and  $E_{\mathfrak{B} \oplus \mathfrak{B}}Y^{-1}$ , respectively. If  $A$  is bounded,  $F, G, F^*$ , and  $G^*$  each have as bounds constants less than 1; if  $A$  is unbounded,  $F, G, F^*$ , and  $G^*$  each have the bound 1, but none of these transformations attains its bound. Finally,*

- (1)  $\mathfrak{N}(I - F) = A_1\mathfrak{B}_A^- = E_{\mathfrak{M}}\mathfrak{B}^-(A),$
- (2)  $\mathfrak{N}(I - F^*) = A_1\mathfrak{B}_A^+ = E_{\mathfrak{M}}\mathfrak{B}^+(A),$
- (3)  $\mathfrak{N}(I - G) = A_1^{-1}\mathfrak{M}_A^- = E_{\mathfrak{B} \oplus \mathfrak{B}}\mathfrak{B}^-(A),$
- (4)  $\mathfrak{N}(I - G^*) = A_1^{-1}\mathfrak{M}_A^+ = E_{\mathfrak{B} \oplus \mathfrak{B}}\mathfrak{B}^+(A).$

That  $F$  and  $G$  are bounded linear with bounds less than or equal to unity is an immediate consequence of their definitions in terms of the isometric transformation  $Y$ . That their ranges are in  $\mathfrak{M}^-$  and  $\mathfrak{B}^-$ , respectively, follows from the equation  $\mathfrak{B}^- + \mathfrak{M}^- = Y(\mathfrak{B}^+ + \mathfrak{M}^+)$  of Theorem 3.8. Since the adjoint from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  of  $Y$  is  $Y^{-1}$  and since  $E_{\mathfrak{B} \oplus \mathfrak{B}}Y^{-1}\mathfrak{M}^-$  is orthogonal to  $\mathfrak{M}^+$ , while  $E_{\mathfrak{M}}Y^{-1}\mathfrak{B}^-$  is orthogonal to  $\mathfrak{B}^+$ , we have at once that  $F^*$  and  $G^*$  are equal on their domains to  $E_{\mathfrak{M}}Y^{-1}$  and  $E_{\mathfrak{B} \oplus \mathfrak{B}}Y^{-1}$ , respectively.

Now consider all elements of  $\mathfrak{B}(A) = \mathfrak{K}(I - Y)$  which are of the form  $\{0, 0, h^+\} - Y\{0, 0, h^+\}$ , where  $h^+$  is in  $\mathfrak{M}^+$ . By Theorem 3.8, every such element can be written in the form  $\{-f^-, if^-, h^+ - h^-\}$ , where  $\{f^-, -if^-\}$  is in  $\mathfrak{B}^-$ ,  $h^-$  in  $\mathfrak{M}^-$ . Conversely, every element of  $\mathfrak{B}^-(A)$  can be written in the form  $\{0, 0, h^+\} - Y\{0, 0, h^+\}$ , where  $h^+$  is in  $\mathfrak{M}^+$ . Furthermore, we have  $h^+ - h^- = -A\{f^-, -if^-\}$ ,  $h^- = Fh^+$ . Let us suppose, then, that the bound of  $F$  is 1. Then we can choose a sequence  $\{h_n^+\}$  in  $\mathfrak{M}^+$  so that  $|h_n^+| = 1$ ,  $|h_n^-| \rightarrow 1$ ,  $n \rightarrow \infty$ , where  $h_n^- = Fh_n^+$ . But, if  $\{f_n^-, -if_n^-, h_n^-\} = Yh_n^+$ , this implies that  $\lim_{n \rightarrow \infty} |\{f_n^-, -if_n^-\}| = 0$ . Consequently, since  $h_n^+ - h_n^- = A\{f_n^-, -if_n^-\}$  and  $\lim_{n \rightarrow \infty} |h_n^+ - h_n^-| = 2^{1/2}$ , it follows that  $A$  is unbounded.

Now let us suppose that  $A$  is unbounded. Then, by Theorem 3.9,  $\mathfrak{B}_A^-$  is a linear manifold dense in  $\mathfrak{B}^-$ ,  $\mathfrak{B}_A^- \neq \mathfrak{B}^-$ . Therefore the bounded linear transformation with domain  $\mathfrak{M}^+$  which is equal on its domain to  $E_{\mathfrak{B} \oplus \mathfrak{B}} Y$  has an unbounded inverse, since otherwise its range, which according to Theorem 3.9 is  $\mathfrak{B}_A^-$ , would be closed. Consequently, we can choose a sequence  $\{\{f_n^-, -if_n^-\}\}$  in  $\mathfrak{B}_A^-$  such that

$$\lim_{n \rightarrow \infty} |\{f_n^-, -if_n^-\}| = 0, \quad |E_{\mathfrak{M}^+ A} \{f_n^-, -if_n^-\}| = 1.$$

Let  $E_{\mathfrak{M}^+ A} \{f_n^-, -if_n^-\} = h_n^+$ . Then

$$|h_n^+|^2 = |\{f_n^-, if_n^-\}|^2 + |Fh_n^+|^2.$$

Thus  $\lim_{n \rightarrow \infty} |Fh_n^+| = 1$ , and  $F$  has the bound 1.

The analogous facts concerning  $G$  are readily established by entirely similar arguments making use of the transformation  $A_1^{-1}$  and the fact that  $A_1^{-1}$  and  $A$  are bounded or unbounded together; we omit the details.

Since  $F^*$  and  $G^*$  evidently have the same bounds as  $F$  and  $G$ , respectively, it follows that  $F^*$  and  $G^*$  have the bound 1 if and only if  $A$  is unbounded.

Now suppose that  $|Fh^+| = |E_{\mathfrak{M}} Yh^+| = |h^+|$  for some  $h^+$  in  $\mathfrak{M}^+$ . Then, since  $Y$  is unitary,  $E_{\mathfrak{B} \oplus \mathfrak{B}} Yh^+ = 0$ , which clearly implies  $h^+ = 0$ . Hence  $F$  never attains the bound 1. Similar arguments serve to establish that  $F^*$ ,  $G$ , and  $G^*$  never attain the bound 1.

The formulas with which the theorem concludes are readily verified and we omit detailed proofs.

**4. Two types of reduction operator.** We have previously had occasion to distinguish between bounded and unbounded reduction operators. For reasons which will be made clear later, it is necessary for us now to distinguish two distinct types of unbounded reduction operator. Furthermore, although it is not necessary, it is clarifying also to distinguish two corresponding types of bounded reduction operator. Hence we introduce the following definition:

DEFINITION 3.5. A reduction operator  $A$  for  $H^*$  is said to be of type I if at least one of the manifolds  $\mathfrak{B}_A^+$ ,  $\mathfrak{B}_A^-$  of Theorem 3.9 contains no closed linear manifold with the dimension number  $\aleph_0$ . Otherwise  $A$  is said to be of type II.

THEOREM 3.12. A bounded reduction operator  $A$  is of type I if and only if one of the manifolds  $\mathfrak{B}^+$ ,  $\mathfrak{B}^-$  is a unitary space.

Theorem 3.12 is an immediate consequence of the fact that when  $A$  is bounded we have, according to Theorem 3.10,  $\mathfrak{B}_A^+ = \mathfrak{B}^+$ ,  $\mathfrak{B}_A^- = \mathfrak{B}^-$ .

In our investigation of the significance of Definition 3.5 for unbounded reduction operators, the following lemma plays an essential role.

LEMMA 3.1. Let  $\mathfrak{H}_1$  and  $\mathfrak{H}_2$  be Hilbert spaces, and let  $T$  be a closed linear transformation with domain dense in  $\mathfrak{H}_1$  and range in  $\mathfrak{H}_2$ . Then a necessary and sufficient condition that  $T$  be totally continuous is that each closed linear manifold in its range have a finite dimension number.

We first take account of the fact that the transformation  $T$  is totally continuous if and only if  $T^*$  is,<sup>†</sup> the existence of  $T^*$  being assured by the hypothesis that  $\mathfrak{D}(T)$  is dense in  $\mathfrak{H}$ . But  $T^* \equiv V(TT^*)^{1/2}$ , where  $V$  is isometric; thus  $T$  is totally continuous if and only if the nonnegative definite self-adjoint transformation  $(TT^*)^{1/2}$  in  $\mathfrak{H}_2$  is totally continuous; furthermore,  $(TT^*)^{1/2}$  has the same range as  $T$ .<sup>‡</sup> Consequently, it is sufficient to prove the lemma for the case that  $\mathfrak{H}_1 = \mathfrak{H}_2 = \mathfrak{H}$  and  $T$  is a nonnegative definite self-adjoint transformation in  $\mathfrak{H}$ ; we now restrict our attention to this case.

But even further simplification is possible. For if  $\mathfrak{N}$  denotes the manifold of zeros of  $T$ ,  $\mathfrak{H} \ominus \mathfrak{N}$  reduces  $T$ ; and, in  $\mathfrak{H} \ominus \mathfrak{N}$ ,  $T$  induces a nonnegative definite self-adjoint transformation  $T_1$  whose inverse exists; moreover,  $T_1$  which has the same range as  $T$ , and is totally continuous if and only if  $T$  is. Consequently, since the trivial case that  $\mathfrak{H} \ominus \mathfrak{N}$  is a unitary space is readily disposed of, it is permissible to assume that the nonnegative definite transformation  $T$  in  $\mathfrak{H}$  has an inverse.

We now prove the necessity of the condition for the specialization of the lemma which we have shown to be equivalent to the general statement. Let us suppose that  $T$  is totally continuous and that  $\mathfrak{M}$  is a closed linear manifold in the range of  $T$ . Let  $\mathfrak{N}$  be the closed linear manifold determined by the set into which  $T^{-1}$  takes  $\mathfrak{M}$ , and observe that since  $T$  is closed and totally continuous, it is bounded and defined throughout  $\mathfrak{H}$ . Hence the transformation  $T_0$  with domain  $\mathfrak{N} \cdot \mathfrak{D}(T)$ , which is equal on its domain to  $T$ , is a bounded linear transformation with domain  $\mathfrak{N}$  and range  $\mathfrak{M}$ . Furthermore  $T_0^{-1}$  exists

<sup>†</sup> Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 100, Theorem 5.

<sup>‡</sup> Murray, Theorem 1.24.

and is closed. Hence, since its domain is the entire space  $\mathfrak{M}$ ,  $T_0^{-1}$  is bounded.† Consequently, if  $\mathfrak{U}$  is a bounded set in  $\mathfrak{M}$ ,  $T_0^{-1}\mathfrak{U}$  is a bounded set in  $\mathfrak{R}$ . But, since  $T$  is totally continuous, this implies that  $\mathfrak{U}$  is a compact set in  $\mathfrak{M}$ . Therefore every bounded set in  $\mathfrak{M}$  is compact, from which it follows that  $\mathfrak{M}$  has a finite dimension number.

[Now let us suppose that no closed linear manifold in  $\mathfrak{R}(T)$  has the dimension number  $\aleph_0$ . Let  $E(\lambda)$  be the resolution of the identity associated with  $T$ , and let  $\{\lambda_n\}$  be a monotone decreasing sequence of positive real numbers convergent to zero and having no member belonging to the point spectrum of  $T$ . Let  $E(\Delta_1) = I - E(\lambda_1)$ ,  $E(\Delta_n) = E(\lambda_{n-1}) - E(\lambda_n)$ ,  $n = 2, 3, 4, \dots$ . Then the range  $\mathfrak{M}_n$  of  $E(\Delta_n)$  reduces  $T$ , and  $T$  induces in  $\mathfrak{M}_n$  a transformation  $T_n$  with bounded inverse. Moreover,  $T_n^{-1}$  is closed and therefore is defined throughout  $\mathfrak{M}_n$ . Hence  $\mathfrak{M}_n$  belongs to the range of  $T$  and, in consequence, is a unitary space. It follows that the points of the spectrum of  $T$  on the interval  $\lambda_{n-1} \leq \lambda \leq \lambda_n$  are finite in number and are all characteristic values (the characteristic values of  $T_n$ ) and that each such characteristic value has finite multiplicity. Thus the characteristic values of  $T$  all have finite multiplicity and can be arranged in a sequence which is bounded and has, as its only possible limit point, the origin, while the continuous spectrum of  $T$  is either empty or contains only the origin. (Here, since  $T^{-1}$  exists, the latter alternative must hold and the origin must be, in fact, a limit point of the spectrum.) Since these spectral properties of  $T$ , deduced from the hypothesis that its range contains no Hilbert spaces, are known to be sufficient for the total continuity of  $T$ ,‡ the lemma stands established.

**THEOREM 3.13.** *Let  $A$  be an unbounded reduction operator, and let  $B_1 \equiv (A_1^* A_1)^{1/2}$ . Let  $\mathfrak{B}$  be the manifold on which  $B_1 = I$ . Then for  $A$  to be of type I it is necessary and sufficient that both of the following conditions be satisfied:*

† Murray, Theorem 1.25.

‡ The theorem which we use here can be stated briefly as follows: If the characteristic values other than zero of a self-adjoint transformation  $T$  each have finite multiplicity and are either finite in number or form a bounded set with zero as its only limit point and if the continuous spectrum of  $T$  contains no point except possibly the origin, then  $T$  is totally continuous. We find no proof of this result in the literature, although the theorem has been stated and used by several writers. It may not be amiss, therefore, to point out here that the theorem follows almost immediately from a theorem of S. Banach (*Théorie des Opérations Linéaires*, Warsaw, 1932, p. 96, Theorem 2). The latter states that the class of all totally continuous transformations  $T$  in a Banach space is closed in the topology defined by setting  $|T_1 - T_2|$  equal to the bound of  $T_1 - T_2$ . Since the spectral properties of  $T$  described in the theorem in question permit us to conclude that  $T$  either has a range with a finite dimension number or is the limit in the sense of the above topology of a sequence of transformations each with that property, and since every linear transformation whose range has a finite dimension number is obviously totally continuous, no further argument is required. Another proof of the theorem, which does not make use of Banach's result, has been communicated to the author by B. Lengyel.

- (1)  $B_1$  induces a totally continuous transformation in the range of  $E_A(1-0)$ ;  
 (2) at least one of the manifolds  $\mathfrak{B}^+ \cdot \mathfrak{P}$  and  $\mathfrak{B}^- \cdot \mathfrak{P}$  is a unitary space.

Since both of the manifolds  $\mathfrak{B}^+ \cdot \mathfrak{P}$  and  $\mathfrak{B}^- \cdot \mathfrak{P}$  belong to  $\mathfrak{D}(B_1) = \mathfrak{D}(A_1)$ , the necessity of (2) is evident. To prove the necessity of (1), we start with the assumption that the transformation  $B_0$ , induced in the range  $\mathfrak{N}$  of  $E_A(1-0)$  by  $B_1$ , is not totally continuous. Then, by Lemma 3.1, there is, in the range of  $B_0$ , a closed linear manifold  $\mathfrak{N}_1$  with the dimension number  $\aleph_0$ . Thus  $\mathfrak{N}_1$  is in the domain of  $B_0^{-1}$ . Therefore, by Theorems 3.5 and 3.6,  $Q\mathfrak{N}_1$  is in the domain of  $B_1$ , where  $Q = iI$  on  $\mathfrak{B}^+$ ,  $Q = -iI$  on  $\mathfrak{B}^-$ . Consequently,  $\mathfrak{N}_1 + Q\mathfrak{N}_1$  is in the domain of  $B_1$ . But since  $\mathfrak{N}_1$  is  $Q$ -symmetric and closed,  $\mathfrak{N}_1 = \mathfrak{N}(I - X)$ ,  $Q\mathfrak{N}_1 = \mathfrak{N}(I + X)$ , where  $X$  is closed and isometric with domain in  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$ , by Theorem 2.2. Hence  $\mathfrak{D}(X)$  and  $\mathfrak{R}(X)$  both belong to the domain of  $B_1$ , and both of these manifolds are Hilbert spaces because  $\mathfrak{N}_1$  is a Hilbert space. Hence, if  $B_0$  is not totally continuous,  $A$  is of type II, and therefore (1) is necessary as we wished to prove.

We prove now the sufficiency of the two conditions. We assume first that  $\mathfrak{B}^+ \cdot \mathfrak{P}$  is a unitary space and denote by  $\mathfrak{N}_0$  the range of the transformation  $B_0$  introduced above. Then, by applying Theorems 3.5 and 3.6, we can resolve the domain of  $A_1$  according to the equation

$$\mathfrak{D}(A_1) = \mathfrak{D}(B_1) = \mathfrak{N} + Q\mathfrak{N}_0 + \mathfrak{B}^+ \cdot \mathfrak{P} + \mathfrak{B}^- \cdot \mathfrak{P}.$$

Now let  $X$  be the isometric transformation associated with  $\mathfrak{N}$  by Theorem 2.2,  $X_0$  the isometric transformation associated with  $\mathfrak{N}_0$ . Then  $\tilde{X}_0 \equiv X$ . Furthermore,  $\mathfrak{D}(X_0)$  contains no Hilbert spaces if  $\mathfrak{N}_0$  contains no Hilbert spaces and thus, according to Lemma 3.1, if  $B_0$  is totally continuous. But  $\mathfrak{D}(A_1) \cdot \mathfrak{B}^+ = (\mathfrak{N} + Q\mathfrak{N}_0) \cdot \mathfrak{B}^+ + \mathfrak{P} \cdot \mathfrak{B}^+$ , and thus  $\mathfrak{B}_A^+ = \mathfrak{D}(A_1) \cdot \mathfrak{B}^+ = \mathfrak{D}(X_0) + \mathfrak{P} \cdot \mathfrak{B}^+$ . Consequently, if  $\mathfrak{P} \cdot \mathfrak{B}^+$  is a unitary space and (1) is satisfied,  $\mathfrak{B}_A^+$  contains no Hilbert spaces. On the other hand, if  $\mathfrak{P} \cdot \mathfrak{B}^-$  contains no closed linear manifold with the dimension number  $\aleph_0$  and (1) is satisfied, an entirely similar argument leads to the conclusion that  $\mathfrak{B}_A^-$  contains no closed linear manifold with the dimension number  $\aleph_0$ . Thus (1) and (2) are sufficient for  $A$  to be of type I and the proof is complete.

**THEOREM 3.14.** *Let  $Y$  have the same meaning as in Theorem 3.8,  $\mathfrak{M}_A^+$  and  $\mathfrak{M}_A^-$  the same meaning as in Theorem 3.10. Then for a reduction operator  $A$  to be of type I, any of the following conditions is necessary and sufficient:*

- (1) At least one of the manifolds  $\mathfrak{M}_A^+$ ,  $\mathfrak{M}_A^-$  contains no closed linear manifold with the dimension number  $\aleph_0$ .  
 (2) Either the transformation with domain  $\mathfrak{M}^+$  which is equal on its domain to  $E_{\mathfrak{B}} - Y$  or the transformation with domain  $\mathfrak{M}^-$  which is equal on its domain to  $E_{\mathfrak{B}} + Y^{-1}$  is totally continuous.

(3) *Either the transformation with domain  $\mathfrak{B}^+$  which is equal on its domain to  $E_{\mathfrak{M}^-}Y$  or the transformation with domain  $\mathfrak{B}^-$  which is equal on its domain to  $E_{\mathfrak{M}^+}Y^{-1}$  is totally continuous.*

By Theorem 3.10, the two transformations defined under (2) have ranges  $\mathfrak{B}_A^-$  and  $\mathfrak{B}_A^+$ , respectively. Thus the necessity and sufficiency of (2) follow at once from Lemma 3.1. Furthermore, since the two transformations defined under (3) have ranges, respectively,  $\mathfrak{M}_A^-$  and  $\mathfrak{M}_A^+$ , the same lemma serves to establish that the conditions (1) and (3) are coextensive. Thus we can complete the proof by showing that (1) is necessary and sufficient, and the latter follows at once when we recall from the proof of Theorem 3.11 that  $\mathfrak{M}_A^- = V\mathfrak{B}_A^+$  and  $\mathfrak{M}_A^+ = V\mathfrak{B}_A^-$ , where  $V$  is isometric.

For use later, we introduce now the following definition:

**DEFINITION 3.6.** *Let  $A$  be a reduction operator of type I, and let  $B_1 \equiv (A_1^*A_1)^{1/2}$ . Let  $\mathfrak{B}$  be the manifold on which  $B_1 = I$ , and let  $j$  and  $k$  be the dimension numbers of  $\mathfrak{B} \cdot \mathfrak{B}^+$  and  $\mathfrak{B} \cdot \mathfrak{B}^-$ , respectively. Then, if  $j$  exceeds  $k$ ,  $(j-k, 0)$  is called the characteristic index of  $A$ . Otherwise,  $(0, k-j)$  is called the characteristic index of  $A$ .*

**THEOREM 3.15.** *Let  $A$  be a bounded reduction operator of type I, and let  $m$  be the dimension number of  $\mathfrak{M}^+$ ,  $n$  the dimension number of  $\mathfrak{M}^-$ . Then, if  $n$  exceeds  $m$ ,  $(n-m, 0)$  is the characteristic index of  $A$ ; otherwise  $(0, m-n)$  is the characteristic index of  $A$ .*

Let  $V$  be the isometric transformation with domain in  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$  associated with the range  $\mathfrak{R}$  of  $E_A(1-0)$  by Theorem 2.2. Then  $\mathfrak{B}^+ \oplus \mathfrak{D}(V) = \mathfrak{B} \cdot \mathfrak{B}^+$ ,  $\mathfrak{B}^- \oplus \mathfrak{R}(V) = \mathfrak{B} \cdot \mathfrak{B}^-$ , by Theorems 3.5 and 2.2, and  $\mathfrak{D}(V)$  and  $\mathfrak{R}(V)$  have the same dimension number. Furthermore, since  $A$  is bounded and of type I, at least one of the manifolds  $\mathfrak{B}^+$ ,  $\mathfrak{B}^-$  is a unitary space. Consequently, since  $\mathfrak{B}^+$  and  $\mathfrak{B}^-$  have the same dimension numbers as  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$ , respectively, we have at once  $n-m = j-k$ , where  $k$  and  $j$  have the same meanings as in Definition 3.6, and the theorem follows.

To conclude this section, we state without formal proof the following theorem:

**THEOREM 3.16.** *Let  $A$  and  $C$  be equivalent reduction operators for  $H^*$ . Then  $A$  and  $C$  are either both bounded or both unbounded and either both of type I with the same characteristic index, or both of type II.*

**5. Self-adjoint extension by means of reduction operators.** A fundamental problem, with which the present work is connected, is the determination of all the self-adjoint extensions in a Hilbert space of a given symmetric transformation  $H$  in  $\mathfrak{S}$ . We define here a quite different type of self-adjoint exten-

sion which is obtained by extending the space  $\mathfrak{S}$  and which is in a certain sense—not the usual one—an “extension” of  $H^*$  or of a transformation whose closure is  $H^*$ . Our construction, as well as the usual one mentioned first, is possible if and only if  $H$  has the deficiency index  $(n, n)$ .

**THEOREM 3.17.** *Let  $H$  be a closed symmetric transformation in a Hilbert space  $\mathfrak{S}$ , and let  $H$  have the deficiency index  $(n, n)$ ,  $(n > 0)$ . Let  $\mathfrak{L}$  be a complex euclidean space with the dimension number  $n$ . Let  $\mathcal{C}$  be the class of all reduction operators  $A$  for  $H^*$  which have the following properties:*

- (1) *the range-space of  $A$  is  $\mathfrak{L} \oplus \mathfrak{L}$ ;*
- (2) *the unitary transformation  $W$  in  $\mathfrak{L} \oplus \mathfrak{L}$ , associated with  $A$  by Definition 1.1, is that which takes  $\{h, k\}$  into  $\{k, -h\}$ .*

*Let  $\mathfrak{S}$  be the class of all self-adjoint transformations  $S$  in  $\mathfrak{S} \oplus \mathfrak{L}$  which have the following properties:*

- (3) *if  $\{f, h\}$  is in the domain of  $S$  and  $S\{f, h\} = \{g, k\}$ , then  $f$  is in the domain of  $H^*$  and  $H^*f = g$ ;*
- (4) *the set of elements  $\{f, H^*f\}$  of  $\mathfrak{B}^*$ , such that  $\{f, h\}$  is in  $\mathfrak{D}(S)$  for some  $h$ , is dense in  $\mathfrak{B}^*$ ;*
- (5) *if  $\{0, h\}$  is in the domain of  $S$ , then  $h = 0$ .*

*Let  $X$  be the transformation which takes  $\mathfrak{S} \oplus \mathfrak{S} \oplus \mathfrak{L} \oplus \mathfrak{L}$  into  $\mathfrak{S} \oplus \mathfrak{L} \oplus \mathfrak{S} \oplus \mathfrak{L}$  according to the equation*

$$X\{f, g, h, k\} = \{f, h, g, k\}.$$

*Then there is a one-to-one correspondence between  $\mathcal{C}$  and  $\mathfrak{S}$  such that, when  $A$  and  $S$  correspond,  $\mathfrak{B}(S) = X\mathfrak{B}(A)$ .*

Let  $A$  be an arbitrary member of  $\mathcal{C}$ , and let  $\mathfrak{N} = X\mathfrak{B}(A)$ . Then  $\mathfrak{N}$  consists of all vectors  $\{f, h, H^*f, k\}$  in  $\mathfrak{S} \oplus \mathfrak{L} \oplus \mathfrak{S} \oplus \mathfrak{L}$  such that  $\{f, H^*f\}$  is in  $\mathfrak{D}(A)$  and  $\{h, k\} = A\{f, H^*f\}$ . Furthermore,  $f$ , and thus  $\{f, h\}$ , determines  $k$  uniquely, since  $f$  determines  $\{h, k\}$  in  $\mathfrak{L} \oplus \mathfrak{L}$  uniquely through the transformation  $A$ . Thus the transformation  $S$  whose domain is the set of all elements  $\{f, h\}$  in  $\mathfrak{S} \oplus \mathfrak{L}$  such that  $f$  is in  $\mathfrak{D}^*$ ,  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$ ,  $\{h, k\} = A\{f, H^*f\}$  for some  $k$  in  $\mathfrak{L}$ , and which is defined by the equation  $\{H^*f, k\} = S\{f, h\}$ , is a one-valued transformation. Furthermore, by Definition 1.1, and condition (2) of the Theorem,  $S$  is self-adjoint, while by definition,  $S$  has the property (3). That  $S$  has the property (4) is a consequence of the fact that  $\mathfrak{D}(A)$  is dense in  $\mathfrak{B}^*$ . That it has the property (5) follows from the fact that  $A$  is one-valued. Thus  $S$  belongs to  $\mathfrak{S}$ .

Now let  $S$  be an arbitrary member of  $\mathfrak{S}$ . Then  $\mathfrak{B}(S)$  consists of a set of vectors of the form  $\{f, h, H^*f, k\}$  in  $\mathfrak{S} \oplus \mathfrak{L} \oplus \mathfrak{S} \oplus \mathfrak{L}$ . Furthermore, according to (5),  $h$  and  $k$  are uniquely determined by  $f$ , and thus by  $\{f, H^*f\}$ . Conse-

quently the transformation  $A$  whose domain is the set of all elements  $\{f, H^*f\}$  of  $\mathfrak{B}^*$  such that for some  $h$  and  $k$  in  $\mathfrak{E}$ ,  $\{f, H, H^*f, k\}$  is in  $\mathfrak{B}(S)$ , and which is defined by the equation  $A\{f, H^*f\} = \{h, k\}$ , is a single-valued transformation. Furthermore, since  $S$  is linear,  $\mathfrak{B}(S)$  is linear; thus  $\mathfrak{B}(A) = X^{-1}\mathfrak{B}(S)$  is linear. Therefore  $A$  is linear. That  $A$  has domain dense in  $\mathfrak{B}^*$  follows from (4); that it satisfies the other conditions of Definition 1.1 is an immediate consequence of the fact that  $S$  is self-adjoint. Since it is readily verified that  $A$  has the properties (1) and (2),  $A$  belongs to  $\mathcal{C}$ .

Finally, since the correspondence is described by the equation  $\mathfrak{B}(S) = X\mathfrak{B}(A)$ , it is clear that it is one-to-one.

We conclude with the observation, readily corroborated with the aid of Theorem 3.3, that if  $H$  has the deficiency index  $(n, n)$ , ( $n \neq 0$ ), every equivalence class of reduction operators for  $H^*$  has in common with the class  $\mathcal{C}$  of Theorem 3.17 a subclass whose cardinal number is that of the continuum.

#### CHAPTER IV. BOUNDARY CONDITIONS

**1. Introduction.** We are now prepared to discuss in detail the fundamental problem proposed in Chapter I, §3, and various questions which arise in connection with it.

As in Chapter III, we consider a fixed but arbitrary symmetric transformation  $H$  in  $\mathfrak{H}$  and the adjoint  $H^*$  of  $H$ . We consider also a fixed but arbitrary reduction operator  $A$  for  $H^*$  and its contraction  $A_1$  with domain  $(\mathfrak{B}^+ + \mathfrak{B}^-) \cdot \mathfrak{D}(A)$ . We preserve the meanings of all the symbols introduced in Definition 3.1 and adopt as standard the symbol  $\mathfrak{M}$  for the range-space of  $A$ , the symbol  $W$  for the unitary transformation in  $\mathfrak{M}$  associated with  $A$  by Definition 1.1, and the symbols  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  for the characteristic manifolds of  $W$  corresponding to the characteristic values  $+i$  and  $-i$ , respectively. Finally, except where otherwise indicated, we continue to reserve the letter  $Q$  for the designation of the transformation of  $\mathfrak{B}^+ + \mathfrak{B}^-$  into itself which is equal to  $iI$  on  $\mathfrak{B}^+$ , to  $-iI$  on  $\mathfrak{B}^-$ .

**2. Boundary conditions involving a bounded reduction operator.** We first dispose of the case that  $\mathfrak{M}$  is a unitary space and of the case that  $\mathfrak{M}$  is a Hilbert space with  $A$  bounded. As we have already pointed out, sufficient information for the solution of our problem in this case is provided by Theorem 3.2.

**THEOREM 4.1.** *Let  $\mathfrak{M}$  be a unitary space, or let  $\mathfrak{M}$  be a Hilbert space with  $A$  bounded. Then there is a one-to-one correspondence between the class of all closed linear symmetric extensions  $S$  of  $H$  and the class of all closed linear  $W$ -symmetric manifolds  $\mathfrak{N}$  in  $\mathfrak{M}$ ;  $S$  and  $\mathfrak{N}$  correspond if  $A\mathfrak{B}(S) = \mathfrak{N}$ , or, in other words, if  $S \equiv H(\mathfrak{N})$  in the sense of Definition 1.2. When  $S$  and  $\mathfrak{N}$  correspond,*

$A\mathfrak{B}(S^*) = (\mathfrak{M} \ominus W\mathfrak{N})$ , and thus  $S^* \equiv H(\mathfrak{M} \ominus W\mathfrak{N})$ . If  $(j, k)$  is the  $W$ -deficiency index of  $\mathfrak{N}$ , then  $(k, j)$  is the deficiency index of  $H(\mathfrak{N})$ . Thus  $H(\mathfrak{N})$  is maximal symmetric if and only if  $\mathfrak{N}$  is maximal  $W$ -symmetric and  $H(\mathfrak{N})$  is self-adjoint if and only if  $\mathfrak{N}$  is hypermaximal  $W$ -symmetric.

Let  $m$  and  $n$  be the dimension numbers of  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$ , respectively. Then, if either  $m$  or  $n$  is zero,  $H$  is maximal symmetric and, if both  $m$  and  $n$  are zero,  $H$  is self-adjoint. If neither  $m$  nor  $n$  is zero, the class of all maximal symmetric extensions  $S$  of  $H$  has the cardinal number of the continuum. If  $m = n < \aleph_0$ , every maximal symmetric extension  $S$  of  $H$  is self-adjoint. If  $m$  exceeds  $n$  ( $m$  does not exceed  $n$ ), every maximal symmetric extension  $S$  of  $H$  has the deficiency index  $(0, m-n)$  ( $(n-m, 0)$ ). If  $m = n = \aleph_0$  and  $p$  is an arbitrary cardinal number on the range  $0 \leq p \leq \aleph_0$ , the class of all maximal symmetric extensions of  $H$  with the deficiency index  $(p, 0)$  ( $(0, p)$ ) has the cardinal number of the continuum.

Excluding the trivial case  $H \equiv H^*$ , we consider first an arbitrary closed symmetric extension  $S$  of  $H$ . Then

$$\mathfrak{B}(S) \subset \mathfrak{B}^* = \mathfrak{D}(A), \quad A\mathfrak{B}(S) = A_1(\mathfrak{B}(S) \ominus \mathfrak{B}).$$

By Theorem 1.3, the latter manifold is  $W$ -symmetric in  $\mathfrak{M}$ . That it is closed is a consequence of the fact that  $A_1^{-1}$  exists and is bounded while  $\mathfrak{B}(S) \ominus \mathfrak{B}$  is closed. On the other hand, if  $\mathfrak{N}$  is an arbitrary closed linear  $W$ -symmetric manifold in  $\mathfrak{M}$ ,  $A_1^{-1}\mathfrak{N}$  is, by a similar argument, a closed linear manifold in  $\mathfrak{B}^* \ominus \mathfrak{B} = \mathfrak{B}^+ + \mathfrak{B}^-$ . Thus  $A_1^{-1}\mathfrak{N} + \mathfrak{B}$  is closed; since, by Theorem 1.5, it is the graph of a linear symmetric extension  $S$  of  $H$ ,  $S$  is closed.

Consider now an arbitrary element  $f$  of the domain of  $S^*$ . Then, since  $WA\mathfrak{B}(S) = W\mathfrak{N}$ , where  $S$  and  $\mathfrak{N}$  correspond,  $Af$  belongs to  $\mathfrak{M} \ominus W\mathfrak{N}$ . On the other hand, since by Theorem 3.2 the range of  $A$  is  $\mathfrak{M}$ , we can determine a solution  $f$  in  $\mathfrak{B}^*$  of the equation  $Af = h$ , for every  $h$  in  $\mathfrak{M} \ominus W\mathfrak{N}$ , and the solution  $f$  obviously belongs to the domain of  $S^*$ . Thus  $A\mathfrak{B}(S^*) = \mathfrak{M} \ominus W\mathfrak{N}$ , or  $S^* \equiv H(\mathfrak{M} \ominus W\mathfrak{N})$ .

Now let  $V$  be the isometric transformation with domain in  $\mathfrak{M}^+$  and range in  $\mathfrak{M}^-$  such that  $\mathfrak{N} = \mathfrak{N}(I - V)$ . Let  $\mathfrak{M}_0 = (\mathfrak{M}^+ \ominus \mathfrak{D}(V)) + (\mathfrak{M}^- \ominus \mathfrak{R}(V))$ . Let  $D$  be the transformation with domain  $\mathfrak{B}(S^*)$  which takes  $f$  into  $E_{\mathfrak{M}_0}Af$ , where  $E_{\mathfrak{M}_0}$  has domain  $\mathfrak{M}$ . Let  $W_0$  be the contraction of  $W$  with domain  $\mathfrak{M}_0$ . We shall show that  $D$  is a reduction operator for  $S^*$ . Invoking Theorem 3.1, we can prove this by showing that the contraction  $D_1$  of  $D$  with domain  $\mathfrak{B}(S^*) \ominus \mathfrak{B}(S)$  is such that  $D_1^* \equiv QD_1^{-1}W_0$ , and this identity is readily shown to be equivalent to  $D_1^{-1} \equiv QD_1^*W_0$ . We shall prove the latter.

We observe first that since  $D_1$  is bounded and has domain  $\mathfrak{B}(S^*) \ominus \mathfrak{B}(S)$ ,  $D_1^*$  exists and has domain  $\mathfrak{M}_0$ . Furthermore, since  $A\mathfrak{B}(S^*) = \mathfrak{M} \ominus W\mathfrak{N}$  and  $A\mathfrak{B}(S) = \mathfrak{N}$ , we see that

$$D_1(\mathfrak{B}(S^*) \ominus \mathfrak{B}(S)) = \mathfrak{M} \ominus (W\mathfrak{N} + \mathfrak{N}) = \mathfrak{M}_0.$$

And, as  $S = H(\mathfrak{N})$ , if  $f$  is in  $\mathfrak{B}(S^*)$  and  $D\{f, S^*f\} = 0$ , then  $A\{f, S^*f\}$  is in  $\mathfrak{N}$  and  $f$  is in  $\mathfrak{D}(S)$ . Consequently, we conclude that  $D_1^{-1}$ , as well as  $D_1^*$ , exists and has domain  $\mathfrak{M}_0$ . Therefore, to establish the identity  $D_1^{-1} = QD_1^*W_0$ , we need only show that  $D_1^{-1}h = QD_1^*W_0h$  for an arbitrary element  $h$  of  $\mathfrak{M}_0$ .

Let  $h$  be such an element, and let  $\{f, S^*f\} = D_1^{-1}h$ . Then  $WA\{f, S^*f\} = W_0h + Wk$ , where  $k$  is in  $\mathfrak{N}$ . Thus, for all  $\{g, S^*g\}$  in  $\mathfrak{B}(S^*) \ominus \mathfrak{B}(S)$ , we obtain

$$(g, S^*f) - (S^*g, f) + (Ag, W_0h + Wk) = 0.$$

But  $Ag = D_1g + r$ , where  $r$  is in  $\mathfrak{N}$ . Thus, because  $(r, W_0h + Wk) = 0$ , and  $(Ag, Wk) = 0$ , we have

$$(g, S^*f) - (S^*g, f) + (D_1g, W_0h) = 0$$

for all  $\{g, S^*g\}$  in  $\mathfrak{B}(S^*) \ominus \mathfrak{B}(S)$ . Consequently, since by Theorem 2.8  $\{S^*f, -f\}$  is in  $\mathfrak{B}(S^*) \ominus \mathfrak{B}$ ,  $W_0h$  is in the domain of  $D_1^*$  and  $Q\{f, S^*f\} = -D_1^*W_0h$ . Therefore, since  $Q^{-1} = -Q$ , we have  $QD_1^*W_0h = D_1^{-1}h$ .

We have thus proved that  $D$  is a reduction operator for  $S^*$ . Furthermore, the characteristic manifolds of  $W_0$  for the characteristic values  $+i$  and  $-i$  are evidently  $\mathfrak{M}^+ \ominus \mathfrak{D}(V)$  and  $\mathfrak{M}^- \ominus \mathfrak{D}(V)$ , respectively; and by definition, the dimension numbers of these manifolds are  $j$  and  $k$ , respectively, where  $(j, k)$  is the  $W$ -deficiency index of  $V$ . Hence, by Theorem 3.7,  $(k, j)$  is the deficiency index of  $S = H(\mathfrak{N})$ . From this it follows at once that  $H(\mathfrak{N})$  is maximal symmetric if and only if  $\mathfrak{N}$  is maximal  $W$ -symmetric, and that  $H(\mathfrak{N})$  is self-adjoint if and only if  $\mathfrak{N}$  is hypermaximal  $W$ -symmetric.

The assertions of the second paragraph of the theorem follow at once from Theorem 3.7 and known results; we state them here because they also follow at once from the assertions of the first paragraph and Theorem 2.5.

We shall find it convenient later to have the following simple facts stated precisely:

**THEOREM 4.2.** *There is a one-to-one correspondence between the class of all linear  $Q$ -symmetric manifolds  $\mathfrak{N}$  in  $\mathfrak{B}^+ + \mathfrak{B}^-$  and the class of all linear symmetric extensions  $S$  of  $H$ ;  $S$  and  $\mathfrak{N}$  correspond if and only if  $\mathfrak{B}(S) = \mathfrak{B} + \mathfrak{N}$  and, when  $S$  and  $\mathfrak{N}$  correspond,*

$$\mathfrak{B}(S^*) = \mathfrak{B} + ((\mathfrak{B}^+ + \mathfrak{B}^- \ominus Q\overline{\mathfrak{N}}).$$

*$S$  is closed if and only if  $\mathfrak{N}$  is a closed linear manifold and, when  $S$  is closed, the deficiency index of  $S$  is the  $Q$ -deficiency index of  $\mathfrak{N}$ .*

In so far as it applies to closed extensions  $S$  of  $H$  and closed manifolds  $\mathfrak{N}$ ,

Theorem 4.2 can be deduced immediately from Theorem 4.1 by setting  $A$  equal to the reduction operator described in Theorem 2.9, the concluding assertion concerning the deficiency index of  $S$  requiring the additional observation that the  $Q$ -deficiency index of  $\mathfrak{R}$  is  $(n, m)$ , where  $(m, n)$  is its  $Q^{-1}$ -deficiency index. That this result can be extended to the more general one stated in Theorem 4.2 is readily verified, and the details may be left to the reader.

Theorems 2.7 and 4.2 contain in essence the basis of the theory of von Neumann in which symmetric extensions of a symmetric transformation  $H$  are determined by means of isometric extensions of its Cayley transform  $(H - iI)(H + iI)^{-1}$ ; to perceive this one has only to note that every isometric transformation  $V$  from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  determines a unique isometric transformation  $X$  from  $\mathfrak{D}^+$  to  $\mathfrak{D}^-$ , and conversely, according to the equation  $V\{f^+, if^+\} = \{Xf^+, -iXf^+\}$ .

**3. Unbounded reduction operators; preliminary questions.** For the case that  $A$  is bounded, we have in the preceding section characterized by means of boundary conditions not only all maximal symmetric extensions of  $H$ , but all closed symmetric extensions of  $H$  and their adjoints. In dealing with the case that  $A$  is unbounded, however, it is not feasible to take in so much territory. Instead, we restrict ourselves almost entirely to the twofold problem stated in Chapter I, §3.

In view of the fact that an unbounded reduction operator is not defined throughout the graph of  $H^*$ , it is natural to inquire first whether or not  $H$  has any maximal symmetric extensions  $S$  such that  $\mathfrak{B}(S)$  belongs to  $\mathfrak{D}(A)$ , and whether or not it has any self-adjoint extensions  $S$  with that property. More generally we may ask if, given an arbitrary cardinal number  $p$  on the range  $0 \leq p \leq \aleph_0$ , there exist symmetric extensions  $S$  of  $H$  with deficiency index  $(0, p)$  ( $(p, 0)$ ) such that  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$ . For the discussion of this question, the classification of reduction operators into types I and II described in §4 of Chapter III is fundamental.

As we have already noted, a bounded reduction operator is of type II if and only if  $H$  has the deficiency index  $(\aleph_0, \aleph_0)$ . Thus, invoking the second paragraph of Theorem 4.1, we can say that a bounded reduction operator  $A$  for  $H^*$  has in its domain the graph of a maximal symmetric extension of  $H$ , with arbitrary preassigned deficiency index  $(0, p)$  ( $(p, 0)$ ), if and only if  $A$  is of type II; and, taking account of Theorem 3.15, we may add that if  $A$  is of type I, every maximal symmetric  $S$  of  $H$  with  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$ , has for its deficiency index the characteristic index of  $A$ . We shall now show that these assertions are valid even when  $A$  is unbounded.

**THEOREM 4.3.** *Let  $A$  be of type II, bounded or unbounded, and let  $p$  be an arbitrary cardinal number on the range  $0 \leq p \leq \aleph_0$ . Then the class of all maximal symmetric extensions  $S$  of  $H$  with the deficiency index  $(0, p)$   $((p, 0))$  such that  $\mathfrak{B}(S)$  belongs to  $\mathfrak{D}(A)$  has the cardinal number  $c$  of the continuum and coincides with the class of all maximal symmetric extensions  $S$  of  $H$  with deficiency index  $(p, 0)$   $((0, p))$  if and only if  $A$  is bounded.*

As we have just observed, if  $A$  is bounded,  $H$  has the deficiency index  $(\aleph_0, \aleph_0)$  and, as we know,  $\mathfrak{D}(A) = \mathfrak{B}(H^*)$ ; thus every maximal symmetric extension  $S$  of  $H$  has its graph in  $\mathfrak{D}(A)$ , and the subclass of extensions  $S$  with deficiency index  $(0, p)$   $((p, 0))$  has the cardinal number of the continuum.

Turning to the case that  $A$  is unbounded, we note first that the class of all symmetric extensions  $S$  of  $H$ , with deficiency index  $(0, p)$   $((p, 0))$  has the cardinal number  $c$ . Hence the cardinal number of the subclass whose members  $S$  satisfy the condition  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$  cannot be greater than  $c$ , so that we need only show that it is at least as great. Moreover, to establish the latter it is sufficient to exhibit a symmetric extension  $T$  of  $H$  with deficiency index  $(\aleph_0, \aleph_0)$  such that  $\mathfrak{B}(T^*) \subset \mathfrak{D}(A)$ ; for the class of all symmetric extensions  $S$  of  $T$ , with deficiency index  $(0, p)$   $((p, 0))$  has the cardinal number  $c$ , and, if  $S$  is such a transformation,  $\mathfrak{B}(S) \subset \mathfrak{B}(T^*) \subset \mathfrak{D}(A)$ .

We now proceed to construct the transformation  $T$ . Again we denote by  $\mathfrak{N}$  the range of  $E_A(1-0)$  and by  $\mathfrak{P}$  the manifold on which  $(A_1^* A_1)^{1/2}$  is equal to  $I$ . Then, if  $\mathfrak{P} \cdot \mathfrak{B}^+$  and  $\mathfrak{P} \cdot \mathfrak{B}^-$  are both Hilbert spaces, the extension  $T$  whose graph is  $\mathfrak{B} + \mathfrak{N}$  is symmetric and has the deficiency index  $(\aleph_0, \aleph_0)$  by Theorem 4.2, since  $\mathfrak{N}$  is  $Q$ -symmetric and has the  $Q$ -deficiency index  $(\aleph_0, \aleph_0)$ ; furthermore  $\mathfrak{B}(T^*) \subset \mathfrak{D}(A)$ .

Now let us assume that at least one of the manifolds  $\mathfrak{P} \cdot \mathfrak{B}^+$  and  $\mathfrak{P} \cdot \mathfrak{B}^-$  is a unitary space. Then, according to Theorem 3.13, the transformation induced in  $\mathfrak{N}$  by  $(A_1^* A_1)^{1/2}$  is not totally continuous and thus, by Lemma 3.1, has in its range a closed linear manifold  $\mathfrak{N}_1$  with the dimension number  $\aleph_0$ . Furthermore,  $\mathfrak{N}_1$  is clearly  $Q$ -symmetric because  $\mathfrak{N}$  is, and  $Q\mathfrak{N}_1$  is in  $\mathfrak{D}(A)$  by Theorems 3.5 and 3.6. Thus  $\mathfrak{N}_1 + Q\mathfrak{N}_1$  is in  $\mathfrak{D}(A)$ . But, applying Theorem 2.2, we see that  $\mathfrak{N}_1 + Q\mathfrak{N}_1 = \mathfrak{D}(X) + \mathfrak{R}(X)$ , where  $\mathfrak{N}_1 = \mathfrak{R}(I - X)$  and  $X$  is a closed isometric transformation with domain in  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$ . Thus  $\mathfrak{D}(X) \subset \mathfrak{D}(A)$  and  $\mathfrak{R}(X) \subset \mathfrak{D}(A)$ . Moreover, since  $\mathfrak{N}_1$  is a Hilbert space,  $\mathfrak{D}(X)$  and  $\mathfrak{R}(X)$  must be Hilbert spaces. Now let  $T$  be the transformation in  $\mathfrak{F}$  whose graph is  $\mathfrak{B} + (\mathfrak{N} \ominus \mathfrak{N}_1)$ . Then by Theorem 4.2, since  $\mathfrak{N} \ominus \mathfrak{N}_1$  is closed linear  $Q$ -symmetric,  $T$  is a closed linear symmetric extension of  $H$ . Furthermore,

$$\mathfrak{B}(T^*) = \mathfrak{B} + (\mathfrak{N} \ominus \mathfrak{N}_1) + \mathfrak{D}(X) + \mathfrak{R}(X) + \mathfrak{P},$$

and therefore  $T$  has the deficiency index  $(\aleph_0, \aleph_0)$  while  $\mathfrak{B}(T^*) \subset \mathfrak{D}(A)$ , as we wished to show.

To complete the proof of the theorem it is necessary only to show that when  $A$  is unbounded, there exist maximal symmetric extensions  $S$  of  $H$ , with deficiency index  $(0, p)$   $((p, 0))$  such that  $\mathfrak{B}(S) \not\subset \mathfrak{D}(A)$ . To construct such an extension  $S$ , we start with a symmetric extension  $S_1$  of  $H$  with the desired deficiency index but such that  $\mathfrak{B}(S_1) \subset \mathfrak{D}(A)$  and denote by  $X_1$  the isometric transformation with domain in  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$  such that  $\mathfrak{B}(S_1) \ominus \mathfrak{B} = \mathfrak{R}(I - X_1)$ . Then either  $\mathfrak{D}(X_1) = \mathfrak{B}^+$  or  $\mathfrak{R}(X_1) = \mathfrak{B}^-$ . Furthermore, since  $\mathfrak{R}(I - X_1) \subset \mathfrak{D}(A)$ ,  $\{f^+, if^+\}$  in  $\mathfrak{D}(X_1)$  belongs to  $\mathfrak{B}_A^+ = \mathfrak{B}^+ \cdot \mathfrak{D}(A)$  if and only if  $X_1\{f^+, if^+\}$  belongs to  $\mathfrak{B}_A^- = \mathfrak{B}^- \cdot \mathfrak{D}(A)$ . Thus we can determine an element  $\{f^+, if^+\}$  of  $\mathfrak{D}(X_1)$  such that neither  $\{f^+, if^+\}$  nor  $X_1\{f^+, if^+\}$  belongs to  $\mathfrak{D}(A)$ . Therefore, since  $(I - X_1)\{f^+, if^+\}$  belongs to  $\mathfrak{D}(A)$ ,  $(I + X_1)\{f^+, if^+\}$  is not contained in  $\mathfrak{D}(A)$ . Consequently, if  $X$  is the transformation with domain  $\mathfrak{D}(X_1)$  which is equal to  $-X_1$  on the manifold determined by  $\{f^+, if^+\}$  and to  $X_1$  on the manifold in  $\mathfrak{D}(X_1)$  perpendicular to  $\{f^+, if^+\}$ , then  $\mathfrak{B} + \mathfrak{R}(I - X)$  is the graph of a symmetric extension  $S$  of  $H$  with the same deficiency index as  $S_1$  and  $\mathfrak{B}(S) \not\subset \mathfrak{D}(A)$ .

The proof of the theorem is thus complete.

We turn now to unbounded operators of type I, beginning with a necessary lemma.

**LEMMA 4.1.** *Let  $\mathfrak{E}$  be a Hilbert space and  $V$  a closed isometric transformation in  $\mathfrak{E}$ , with deficiency index  $(j, k)$ ,  $j \neq k$ . Then the range of  $I - V$  contains a Hilbert space—that is, a closed linear manifold with the dimension number  $\aleph_0$ .*

We denote by  $\mathfrak{E}_1$  the manifold on which  $V = I$ . Then it is readily shown that  $\mathfrak{R}(I - V) \subseteq \mathfrak{E} \ominus \mathfrak{E}_1$ . Now let  $V_1$  be the transformation with domain  $(\mathfrak{E} \ominus \mathfrak{E}_1) \cdot \mathfrak{D}(V)$  which is equal on its domain to  $V$ . Then  $\mathfrak{R}(V_1) \subset \mathfrak{E} \ominus \mathfrak{E}_1$ . Thus, since  $\mathfrak{E}_1 \subset \mathfrak{D}(V)$  and  $\mathfrak{E}_1 \subset \mathfrak{R}(V)$ ,  $V_1$  is an isometric transformation in  $\mathfrak{E} \ominus \mathfrak{E}_1$  with the deficiency index  $(j, k)$ . Moreover, since  $j \neq k$ ,  $\mathfrak{E} \ominus \mathfrak{E}_1$  is clearly a Hilbert space.

We now determine in  $\mathfrak{E} \ominus \mathfrak{E}_1$  a maximal symmetric extension  $V_2$  of  $V_1$  such that  $(I - V_2)^{-1}$  exists; for this construction it is necessary only to choose a maximal isometric transformation  $V_3$  from  $(\mathfrak{E} \ominus \mathfrak{E}_1) \ominus \mathfrak{D}(V_1)$  to  $(\mathfrak{E} \ominus \mathfrak{E}_1) \ominus \mathfrak{R}(V_1)$  such that  $V_3 \neq I$  at any point of its domain and then to define  $V_2 = V_1$  on  $\mathfrak{D}(V_1)$ ,  $V_2 = V_3$  on  $\mathfrak{D}(V_3)$ . Evidently the deficiency index of  $V_2$  in  $\mathfrak{E} \ominus \mathfrak{E}_1$  is either  $(j - k, 0)$  or  $(0, k - j)$ , according as  $j$  does or does not exceed  $k$ .

Next we observe that  $I - V_2$  has range dense in  $\mathfrak{E} \ominus \mathfrak{E}_1$ . For, if  $(f - V_2f, g) = 0$  for all  $f$  in  $\mathfrak{D}(V_2)$  and some  $g$  in  $\mathfrak{E} \ominus \mathfrak{E}_1$ , we have  $(f, g) = (V_2f, g)$  and this is

readily shown to imply  $g = V_2 g$ , which in turn implies  $g = 0$ . Consequently  $S_2 = (I + V_2)(I - V_2)^{-1}$  is a maximal symmetric transformation in  $\mathfrak{L} \ominus \mathfrak{L}_1$  and

$$S_2 \supseteq S_1 = (I + V_1)(I - V_1)^{-1}. \dagger$$

Furthermore, applying Theorems 2.7 and 2.3, we have  $\mathfrak{B}(S_2) \ominus \mathfrak{B}(S_1) = \mathfrak{B}(S_3)$ , where  $S_3 = (I + V_3)(I - V_3)^{-1}$  and  $V_3$  has the same meaning as above. Moreover,  $\mathfrak{B}(S_3)$  is a unitary space, because its dimension number is the minimum of  $j$  and  $k$  and, since  $j \neq k$ , the minimum of the two is finite.

Now let us suppose that  $\mathfrak{D}(S_2)$  contains a Hilbert space  $\mathfrak{L}_2$  and denote by  $S_4$  the contraction of  $S_2$ , with domain  $\mathfrak{L}_2$ . Then  $S_4$  is evidently closed and  $\mathfrak{B}(S_4)$  is thus a closed linear manifold in  $\mathfrak{B}(S_2)$ . Let  $T$  be the transformation with domain  $\mathfrak{B}(S_4)$  and range in  $\mathfrak{B}(S_3)$  which takes each element of its domain into the projection of that element on  $\mathfrak{B}(S_3)$ , let  $\mathfrak{N}$  be the manifold of zeros of  $T$ , and let  $\mathfrak{N}_1 = \mathfrak{B}(S_4) \ominus \mathfrak{N}$ . Then the dimension number of  $\mathfrak{N}_1$  is clearly not greater than that of  $\mathfrak{B}(S_3)$ . Thus, since  $\mathfrak{B}(S_4)$  is a Hilbert space and  $\mathfrak{B}(S_3)$  a unitary space,  $\mathfrak{N}$  must be a Hilbert space. Moreover,  $\mathfrak{N} = \mathfrak{B}(S_4) \cdot \mathfrak{B}(S_1)$  and is evidently the graph of a closed linear transformation  $S_5$ ,  $S_5 \subseteq S_1$ ,  $S_5 \subseteq S_4$ . But  $S_4$ , having domain the closed space  $\mathfrak{L}_2$  and being closed itself, is therefore bounded. ‡ Hence  $S_5$  is bounded; therefore, since  $S_5$  is also closed,  $\mathfrak{D}(S_5)$  is a closed linear manifold. Furthermore,  $\mathfrak{D}(S_5)$  has the dimension number  $\aleph_0$ , because  $\mathfrak{B}(S_5)$  has that dimension number, and

$$\mathfrak{D}(S_5) \subset \mathfrak{D}(S_1) = \mathfrak{N}(I - V_1) = \mathfrak{N}(I - V).$$

In consequence of the result just obtained, it is necessary for the completion of the proof only to show that the domain of the maximal symmetric transformation  $S_2$  contains a Hilbert space  $\mathfrak{L}_2$ . To prove this we first recall that since  $S_2$  is maximal symmetric and not self-adjoint, there is a Hilbert space  $\mathfrak{L}_3$  which reduces  $S_2$  and in which  $S_2$  induces an elementary symmetric transformation. § Therefore, by definition, in  $\mathfrak{L}_3$ ,  $S_2 = (I + X)(I - X)^{-1}$ , where  $X$  is defined in  $\mathfrak{L}_3$ , in terms of some complete orthonormal set  $\{\phi_n\}$ ,  $n = 1, 2, 3, \dots$ , by the equations  $X\phi_n = \phi_{n+1}$ ,  $n = 1, 2, 3, \dots$ , or by the equations  $X\phi_n = \phi_{n-1}$ ,  $n = 2, 3, \dots$ . || From this fact it is evident that the orthonormal set  $\{(\phi_n - \phi_{n-1})/2\}$ ,  $n = 2, 3, \dots$ , determines a Hilbert space  $\mathfrak{L}_2$  which belongs to  $\mathfrak{N}(I - X)$  and thus to  $\mathfrak{D}(S_2)$ . Consequently the lemma is true as stated.

† Stone, Theorems 9.1 and 9.2.

‡ Murray, Theorem 1.25.

§ Stone, Theorem 9.10.

|| Stone, Theorem 9.9 and Definition 9.6.

**THEOREM 4.4.** *Let  $A$  be a reduction operator of type I. Then, if  $H$  is not maximal symmetric, the class of all maximal symmetric extensions  $S$  of  $H$  such that  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$  has the cardinal number  $c$  of the continuum and coincides with the class of all maximal symmetric extensions  $S$  of  $H$  if and only if  $A$  is bounded. If  $S$  is a maximal symmetric extension of  $H$  such that  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$ , then the deficiency index of  $S$  is the characteristic index of  $A$ . If  $A$  is unbounded, there exist maximal symmetric extensions  $S$  of  $H$  with this deficiency index such that  $\mathfrak{B}(S) \not\subset \mathfrak{D}(A)$ .*

If  $A$  is bounded,  $\mathfrak{D}(A) = \mathfrak{B}(H^*)$ , and every maximal symmetric extension  $S$  of  $H$  has its graph in  $\mathfrak{D}(A)$ ; furthermore, if  $H$  is not maximal, the cardinal number of the class of all such extensions  $S$  is the cardinal number  $c$  of the continuum, by Theorem 4.1. That every such extension has for its deficiency index the characteristic index of  $A$  follows from Theorems 3.15 and 4.1.

Now let  $A$  be unbounded. Invoking an argument like that used in the proof of Theorem 4.3, we can prove that the class defined in the theorem has the cardinal number  $c$  by exhibiting a subclass with that cardinal number. Moreover, for this purpose it is clearly sufficient to construct a closed symmetric extension  $S_1$  of  $H$ , which is not maximal and which satisfies the condition  $\mathfrak{B}(S_1^*) \subset \mathfrak{D}(A)$ .

To construct  $S_1$ , we start with the transformation  $S_2$  whose graph is  $\mathfrak{N} + \mathfrak{B}$  where  $\mathfrak{N}$  is the range of  $E_A(1-0)$ . Then by Theorem 4.2, if  $\mathfrak{P}$  is the manifold on which  $(A_1^* A_1)^{1/2} = I$ ,  $S_2$  is a closed symmetric extension of  $H$  with the deficiency index  $(j, k)$ , where  $j$  and  $k$  are the dimension numbers of  $\mathfrak{P} \cdot \mathfrak{B}^+$  and  $\mathfrak{P} \cdot \mathfrak{B}^-$ , respectively. Furthermore, since  $\mathfrak{B}_A^+$  contains  $\mathfrak{P} \cdot \mathfrak{B}^+$  and is dense in  $\mathfrak{B}^+$ ,  $(\mathfrak{B}^+ \ominus \mathfrak{B}^+ \cdot \mathfrak{P}) \cdot \mathfrak{B}_A^+$  contains an element  $\{f^+, if^+\} \neq 0$ . Now let  $X_2$  be the isometric transformation with domain  $(\mathfrak{B}^+ \ominus \mathfrak{B}^+ \cdot \mathfrak{P})$  and range  $(\mathfrak{B}^- \ominus \mathfrak{B}^- \cdot \mathfrak{P})$  such that  $\mathfrak{N} = \mathfrak{N}(I - X_2)$ , and let  $X_1$  be the contraction of  $X_2$  whose domain is the set of elements of  $\mathfrak{D}(X_2)$  orthogonal to  $\{f^+, if^+\}$ . Then by Theorem 4.2,  $\mathfrak{B} + \mathfrak{N}(I - X_1)$  is the graph of a symmetric extension  $S_1$  of  $S$  and

$$\mathfrak{B}(S_1^*) = \mathfrak{B} + \mathfrak{N}(I - X_1) + \mathfrak{B}^+ \cdot \mathfrak{P} + \mathfrak{B}^- \cdot \mathfrak{P} + \mathfrak{B}_I^+ + \mathfrak{B}_I^-,$$

where  $\mathfrak{B}_I^+$  is the linear manifold determined by  $\{f^+, if^+\}$ ,  $\mathfrak{B}_I^-$  the linear manifold determined by  $X_2\{f^+, if^+\}$ . Moreover, since  $\{f^+, if^+\}$  and  $(I - X)\{f^+, if^+\}$  belong to  $\mathfrak{D}(A)$ ,  $X_2\{f^+, if^+\}$  does also. Thus  $\mathfrak{B}(S_1^*) \subset \mathfrak{D}(A)$ . Therefore, since  $S_1$  clearly has the deficiency index  $(j+1, k+1)$ , the class of maximal symmetric extensions  $S$  of  $H$  such that  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$  has the cardinal number  $c$ .

To prove that every such extension has the deficiency index stated in the theorem we note first that every maximal symmetric extension  $S$ , of the transformation  $S_1$  just constructed, has for its deficiency index the character-

istic index of  $A$  (that is, either  $(j-k, 0)$  or  $(0, k-j)$  according as  $j$  does or does not exceed  $k$ ). Let  $S$  be such an extension, and let  $\mathfrak{B}(S) \ominus \mathfrak{B} = \mathfrak{R}(I-X)$ , where  $X$  is the closed isometric transformation with domain in  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$ , which exists in accordance with Theorems 4.2 and 2.2. Let  $T$  be another maximal symmetric extension of  $H$  such that  $\mathfrak{D}(T) \subset \mathfrak{D}(A)$ , and let  $\mathfrak{B}(T) \ominus \mathfrak{B} = \mathfrak{R}(I-Y)$  where  $Y$  is closed isometric with domain in  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$ . Then, since both  $\mathfrak{R}(I-X)$  and  $\mathfrak{R}(I-Y)$  belong to  $\mathfrak{D}(A)$ ,  $\mathfrak{R}(X-Y)$  belongs to  $\mathfrak{D}(A)$ . Moreover,

$$\mathfrak{R}(I-X) = \mathfrak{R}((I-X)X^{-1}) = \mathfrak{R}(I-X^{-1})$$

and, similarly,

$$\mathfrak{R}(I-Y) = \mathfrak{R}(I-Y^{-1}),$$

so that  $\mathfrak{R}(X^{-1}-Y^{-1})$  belongs to  $\mathfrak{D}(A)$  also. Thus  $\mathfrak{R}(X-Y)$  is in  $\mathfrak{B}_A^-$  and  $\mathfrak{R}(X^{-1}-Y^{-1})$  is in  $\mathfrak{B}_A^+$ . Furthermore,

$$\mathfrak{R}(X-Y) = \mathfrak{R}((X-Y)Y^{-1}) = \mathfrak{R}(I-XY^{-1})$$

and

$$\mathfrak{R}(X^{-1}-Y^{-1}) = \mathfrak{R}((X^{-1}-Y^{-1})X) = \mathfrak{R}(I-Y^{-1}X).$$

Now let us suppose that the deficiency index of  $S$  is  $(0, p)$  and that of  $T$  is  $(q, 0)$ . Then  $XY^{-1}$  is a maximal isometric transformation in  $\mathfrak{B}^-$  with deficiency index  $(0, p+q)$  and  $Y^{-1}X$  is a maximal symmetric transformation in  $\mathfrak{B}^+$  with the same deficiency index. Hence, since  $\mathfrak{R}(I-XY^{-1}) \subseteq \mathfrak{B}_A^-$  and  $\mathfrak{R}(I-Y^{-1}X) \subseteq \mathfrak{B}_A^+$ , it follows, from Lemma 4.1 and the hypothesis that  $A$  is of type I, that  $p=q=0$ . Moreover, by an entirely similar argument, the assumption that  $S$  has the deficiency index  $(p, 0)$  and  $T$  the deficiency index  $(0, q)$  leads to the same conclusion.

Next let us assume that  $S$  has the deficiency index  $(p, 0)$ ,  $T$  the deficiency index  $(q, 0)$ ; then  $Y^{-1}X$  has the deficiency index  $(p, q)$ . Let us assume  $p \neq q$ . Then, by Lemma 4.1, since  $\mathfrak{B}_A^+ \supseteq \mathfrak{R}(I-Y^{-1}X)$ ,  $\mathfrak{B}_A^+$  contains a Hilbert space  $\mathfrak{B}_2^+$ . Furthermore, since either  $p$  or  $q$  is finite, either  $\mathfrak{B}^+ \ominus \mathfrak{D}(X)$  or  $\mathfrak{B}^+ \ominus \mathfrak{D}(Y)$  is a unitary space. Therefore, by an argument like one used in the proof of Lemma 4.1, either  $\mathfrak{D}(X) \cdot \mathfrak{B}_2^+$  or  $\mathfrak{D}(Y) \cdot \mathfrak{B}_2^+$  is a Hilbert space  $\mathfrak{B}_3^+$ . Moreover, since  $\mathfrak{R}(I-X)$ ,  $\mathfrak{R}(I-Y)$ , and  $\mathfrak{B}_2^+$  belong to  $\mathfrak{D}(A)$ , either  $X$  or  $Y$  takes  $\mathfrak{B}_3^+$  into a Hilbert space in  $\mathfrak{B}_A^-$ . But this is evidently a contradiction of the hypothesis that  $A$  is of type I, and we conclude in consequence that the assumption  $p \neq q$  is untenable.

Since the case that  $S$  and  $T$  have deficiency indices  $(0, p)$  and  $(0, q)$ , respectively, can evidently be handled in a manner entirely similar to the above, we conclude that  $T$  has the same deficiency index as  $S$ .

To complete the proof, it is now only necessary to show that when  $A$  is unbounded there exists a maximal symmetric extension  $S$  of  $H$ , whose deficiency index is the characteristic index of  $A$  and whose graph is not in  $\mathfrak{D}(A)$ . A method for the proof of this, however, has already been used in proving a similar assertion in Theorem 4.2; we leave the details here to the reader.

We now prove a theorem which establishes the significance of the second part of our fundamental problem, stated in Chapter I, §3.

**THEOREM 4.5.** *Let  $A$  be unbounded, and let  $p$  be an arbitrary cardinal number on the range  $0 \leq p \leq \aleph_0$ . Let  $S$  be an arbitrary maximal symmetric extension of  $H$ . Then  $\mathfrak{B}(S) \cdot \mathfrak{D}(A)$  is the graph of a symmetric extension  $S_A$  of  $H$ . The class of all maximal symmetric extensions  $S$  of  $H$ , with the deficiency index  $(0, p)$   $((p, 0))$ , such that  $S_A \neq S$ ,  $\tilde{S}_A \equiv S$ , has the cardinal number  $c$  of the continuum.*

The first assertion of the theorem is obvious. To prove the second, we observe first that, since the linear manifold  $\mathfrak{B}_A^+ = \mathfrak{D}(A) \cdot \mathfrak{B}^+$  is dense in  $\mathfrak{B}^+$  and the linear manifold  $\mathfrak{B}_A^- = \mathfrak{D}(A) \cdot \mathfrak{B}^-$  is dense in  $\mathfrak{B}^-$ , we can choose in  $\mathfrak{B}_A^+$  an orthonormal set  $\{f_n^+, if_n^+\}$  complete in  $\mathfrak{B}^+$  and in  $\mathfrak{B}_A^-$  an orthonormal set  $\{f_n^-, -if_n^-\}$  complete in  $\mathfrak{B}^-$ . Hence if  $V$  is the closed isometric transformation defined in terms of the two sets by the equation

$$V\{f_n^+, if_n^+\} = \{f_{n+p}^-, -if_{n+p}^-\}, \quad (V\{f_{n+p}^+, if_{n+p}^+\} = \{f_n^-, -if_n^-\}),$$

$\mathfrak{B} + \mathfrak{R}(I - V)$  is clearly the graph of a maximal symmetric extension  $S$  of  $H$ , with deficiency index  $(0, p)$   $((p, 0))$ , such that the transformation  $S_A$  defined in the theorem has  $S$  for its closure. Moreover, the class of all transformations  $S$  which can be constructed in this way evidently has the cardinal number  $c$ . Finally, if  $S \equiv S_A$ , the transformation  $T$  whose graph is  $\mathfrak{B} + \mathfrak{R}(I + V)$  obviously has the same deficiency index as  $S$  and satisfies the conditions  $T_A \neq T$ ,  $\tilde{T}_A \equiv T$ . Thus, again taking account of the fact that the class of all maximal symmetric extensions of  $H$  with deficiency index  $(0, p)$   $((p, 0))$  has the cardinal number  $c$ , we conclude that the class defined in the theorem has the same cardinal number.

It is to be emphasized that when  $A$  is unbounded, the maximal symmetric extensions of  $H$  described in Theorems 4.3, 4.4, and 4.5, do not exhaust the class of all maximal symmetric extensions of  $H$ . We now indicate briefly the wide range of other possibilities.

We begin by considering an arbitrary maximal symmetric extension  $S_1$  of  $H$  such that  $\mathfrak{B}(S_1) \subset \mathfrak{D}(A)$ . For simplicity, let us assume that the deficiency index of  $S_1$  is  $(0, p)$ . Then  $\mathfrak{B}(S_1) \ominus \mathfrak{B} = \mathfrak{R}(I - X)$ , where  $X$  is isometric with domain  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$ . Now let  $T$  be any self-adjoint transformation in  $\mathfrak{B}^+$ , and let  $V \equiv (T - iT)(T + iT)^{-1}$ . Then  $V$  is unitary in  $\mathfrak{B}^+$ , and  $V \equiv XV$

is isometric with domain  $\mathfrak{B}^+$  and range in  $\mathfrak{B}^-$ . Thus, by Theorems 2.2 and 4.2,  $\mathfrak{B} + \mathfrak{R}(I - Y) = \mathfrak{B} + \mathfrak{R}(I - Y^{-1})$  is the graph of a maximal symmetric extension  $S$  of  $H$ . Furthermore, for every  $h$  in  $\mathfrak{R}(Y) = \mathfrak{R}(X)$ , we have

$$h - Y^{-1}h = (h - X^{-1}h) + (X^{-1}h - V^{-1}X^{-1}h).$$

Consequently since  $\mathfrak{R}(I - X^{-1})$  is in  $\mathfrak{D}(A)$  by hypothesis,  $h - Y^{-1}h$  is in  $\mathfrak{D}(A)$  if and only if  $(X^{-1}h - V^{-1}X^{-1}h)$  is in  $\mathfrak{D}(A)$  and thus in  $\mathfrak{B}_A^+ = \mathfrak{D}(A) \cdot \mathfrak{B}^+$ . But  $\mathfrak{R}(X^{-1}) = \mathfrak{B}^+$  and  $\mathfrak{R}(I - V^{-1}) = \mathfrak{D}(T)$ . Therefore  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$  if and only if  $\mathfrak{D}(T) \subseteq \mathfrak{B}_A^+$  and  $\mathfrak{B}(S) \cdot \mathfrak{D}(A)$  is dense in  $\mathfrak{B}(S)$  if and only if  $T$  is the closure of a transformation  $T_1$  with domain in  $\mathfrak{B}_A^+$ .

Evidently, when  $A$  is unbounded  $T$  can be chosen in a wide variety of ways so that neither of these conditions is satisfied; in particular, we can proceed as follows. According to Theorem 3.10,  $\mathfrak{B}_A^+$  is the range of a closed transformation  $M$  with domain  $\mathfrak{M}^-$ , such that  $M^{-1}$  exists. Hence  $T_0 = (M^{-1}(M^*)^{-1})^{1/2}$  is a self-adjoint transformation with domain  $\mathfrak{B}_A^+$ . Furthermore, if  $A$  is unbounded,  $\mathfrak{B}_A^+ \neq \mathfrak{B}^+$  and  $T_0$  is unbounded. Consequently there exists a unitary transformation  $U$  in  $\mathfrak{B}^+$  such that  $\mathfrak{D}(UT_0U^{-1}) \cdot \mathfrak{D}(T_0) = \mathfrak{D}$ . Let  $T \equiv UT_0U^{-1}$ .† Then, if  $S$  is the maximal symmetric extension of  $H$  determined by  $T$  and the extension  $S_1$  of  $H$  in the manner described immediately above, we clearly have  $\mathfrak{B}(S) \cdot \mathfrak{D}(A) = \mathfrak{B}$ . Thus, since an entirely similar procedure is possible if the transformation  $S_1$  has the deficiency index  $(p, 0)$ , we can state the following theorem:

**THEOREM 4.6.** *If  $A$  is an unbounded reduction operator, there exist maximal symmetric extensions  $S$  of  $H$  such that  $\mathfrak{B}(S)$  and  $\mathfrak{D}(A)$  intersect only on the graph of  $H$ .*

Further light on the pathological aspects of the theory of unbounded reduction operators is provided by the following theorem which was communicated to the author by J. von Neumann: If  $H$  has the deficiency index  $(\aleph_0, \aleph_0)$ , there exist unbounded reduction operators  $A$  and  $C$  for  $H^*$  such that  $\mathfrak{D}(A) \cdot \mathfrak{D}(C) = \mathfrak{B}$ . In fact, it can be shown that the operator  $A$  can be an arbitrary unbounded reduction operator for  $H^*$  and  $C$  determined so that  $\mathfrak{D}(A) \cdot \mathfrak{D}(C) = \mathfrak{B}$ . Thus Theorem 4.6 can be obtained as a consequence of Theorems 4.3 and 4.4.

**4. Boundary conditions involving an arbitrary reduction operator.** We now return to the consideration of the maximal symmetric extensions of  $H$  described in Theorems 4.3, 4.4, and 4.5. According to Theorems 1.2 and 1.3, each such extension  $S$ , or a transformation  $S_A$  whose closure is  $S$ , can be de-

† Von Neumann, *Journal für die reine und angewandte Mathematik* (Crelle), vol. 161 (1929), pp. 208-236; Theorem 17.

finer by means of a boundary condition of the kind described in Definition 1.2. In order to describe what special properties of the boundary condition are equivalent to the maximal property of the transformation  $S$ , it is convenient to vary slightly our previous notation.

**DEFINITION 4.1.** Let  $\mathfrak{R}$  be a linear  $W$ -symmetric manifold in  $\mathfrak{M}$ , and let  $V$  be the isometric transformation with domain in  $\mathfrak{M}^+$  and range in  $\mathfrak{M}^-$ , such that  $\mathfrak{R} = \mathfrak{R}(I - V)$  associated with  $\mathfrak{R}$  by Theorem 2.2. Then  $H(V)$  denotes the same transformation as  $H(\mathfrak{R})$ , where  $H(\mathfrak{R})$  is the operator defined in Definition 1.2.

**THEOREM 4.7.** The transformation  $H(V)$  of Definition 4.1 is a linear symmetric extension of  $H$ . If  $S$  is an arbitrary linear symmetric extension of  $H$  such that  $\mathfrak{B}(S) \subseteq \mathfrak{D}(A)$ , there exists an isometric transformation  $V$  with domain in  $\mathfrak{M}^+$  and range in  $\mathfrak{M}^-$  such that  $S \equiv H(V)$ .

Theorem 4.7 is essentially only a restatement of Theorem 1.5.

We emphasize that  $H(V_1) \equiv H(V_2)$  does not imply  $V_1 \equiv V_2$ . For example,  $V_1$  can be the transformation whose domain contains only the element 0 of  $\mathfrak{M}^+$ , while  $V_2$  has for its domain the linear manifold determined by an element  $h^+$  of  $\mathfrak{M}^+$  such that  $h^+ - V_2 h^+$  is not in  $\mathfrak{R}(A)$ . We then have  $H \equiv H(V_1) \equiv H(V_2)$ .

**THEOREM 4.8.** Let  $F$  be the transformation defined in Theorem 3.11, and let  $\mathfrak{M}_A^+$  and  $\mathfrak{M}_A^-$  have the same meanings as in Theorem 3.8. Then a necessary and sufficient condition that the transformation  $H(V)$  of Definition 4.1 be maximal symmetric with deficiency index  $(0, p)$  ( $(p, 0)$ ) is that  $\mathfrak{R}(V - F) \supseteq \mathfrak{M}_A^-$  ( $\mathfrak{R}(V^{-1} - F^*) \supseteq \mathfrak{M}_A^+$ ). Thus  $H(V)$  is self-adjoint if and only if  $\mathfrak{R}(V - F) \supseteq \mathfrak{M}_A^-$  and  $\mathfrak{R}(V^{-1} - F^*) \supseteq \mathfrak{M}_A^+$ .

If  $S$  is an arbitrary maximal symmetric extension of  $H$  and  $\mathfrak{B}(S) \subseteq \mathfrak{D}(A)$ , there exists one and only one closed isometric transformation  $V$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that the boundary condition  $Af \in \mathfrak{R}(I - V)$  is nondegenerate and such that  $S \equiv H(V)$ .

Our initial step in the proof of the first portion of the theorem is purely formal, leading to an equation on which the proof is based. We consider an arbitrary element  $\{f, H^*f, Af\}$  of  $\mathfrak{B}(A)$  such that  $Af \in \mathfrak{R}(I - V)$  and set

$$Af = h^+ - Vh^+,$$

$$\{f, H(V)f\} = \{f^+, if^+\} - X\{f^+, if^+\} + \{f_0, Hf_0\},$$

$X$  being the unique isometric transformation from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that  $\mathfrak{R}(I - X) = \mathfrak{B}(S) \cdot (\mathfrak{B}^+ + \mathfrak{B}^-)$ , whose existence is assured by Theorems 4.2 and 2.2. We then apply Theorem 3.9 to write

$$\{f, H(V)f, Af\} = \{f_-, H^*f_-, Af_-\} - \{f^-, -if^-, Af^-\} + \{f_0, Hf_0, 0\},$$

where the first component on the right is in  $\mathfrak{B}_-(A)$ , the second in  $\mathfrak{B}^-(A)$ . Recalling that  $Af_-$  is in  $\mathfrak{M}^-$ , and comparing the first and third equations above, we have

$$-Af_- = h^+ - Fh^+, \quad Af_- = -(V - F)h^+.$$

Likewise, comparing the second and third, we obtain

$$\{f_-, H^*f_-\} = \{f^+, if^+\} - G\{f^+, if^+\}, \quad \{f_-, -if^-\} = (X - G)\{f^+, if^+\},$$

where  $G$  has the same meaning as in Theorem 3.11. Consequently,

$$(4.1) \quad A_1(I - G)\{f^+, if^+\} = -(V - F)h^+$$

for all  $\{f^+, if^+\}$  in  $\mathfrak{D}(X)$  or, to put it differently, for all  $h^+$  in  $\mathfrak{D}(V)$  such that  $h^+ - Vh^+ \in \mathfrak{R}(A)$ .

Thus from (4.1) we have

$$(4.2) \quad A_1(I - G)\mathfrak{D}(X) \subseteq \mathfrak{R}(V - F).$$

Furthermore  $A_1^{-1}$  and  $(I - G)^{-1}$  exist; so we may write

$$(4.3) \quad \mathfrak{D}(X) \subseteq (I - G)^{-1}A_1^{-1}[\mathfrak{R}(V - F) \cdot \mathfrak{R}(A)].$$

But, if  $(V - F)h^+$  is in  $\mathfrak{R}(A)$ , then

$$(I - V)h^+ = (I - F)h^+ - (V - F)h^+$$

is also, since  $(I - F)h^+$  is in  $\mathfrak{R}(A)$  by definition of  $F$ . Therefore (4.3) becomes

$$(4.4) \quad \mathfrak{D}(X) = (I - G)^{-1}A_1^{-1}[\mathfrak{R}(V - F) \cdot \mathfrak{M}_A^-],$$

where we have made use of the relations  $\mathfrak{R}(V - F) \subseteq \mathfrak{M}^-$ ,  $\mathfrak{M}_A^- = \mathfrak{M}^- \cdot \mathfrak{R}(A)$ ; in addition, we can evidently replace (4.2) by

$$(4.5) \quad A_1(I - G)\mathfrak{D}(X) = \mathfrak{R}(V - F) \cdot \mathfrak{M}_A^-.$$

But, by Theorem 3.11, (3),  $A_1(I - G)\mathfrak{B}^+ = \mathfrak{M}_A^-$ , and besides, it is clear that  $\mathfrak{B}^+ = (I - G)^{-1}A_1^{-1}\mathfrak{M}_A^-$ . Consequently we conclude from equations (4.4) and (4.5) that  $\mathfrak{D}(X) = \mathfrak{B}^+$  if and only if  $\mathfrak{R}(V - F) \cdot \mathfrak{M}_A^- = \mathfrak{M}_A^-$ ; that is to say, if and only if  $\mathfrak{R}(V - F) \supseteq \mathfrak{M}_A^-$ . Moreover,  $\mathfrak{D}(X) = \mathfrak{B}^+$  if and only if  $\mathfrak{R}(I - X)$  is maximal  $Q$ -symmetric with  $Q$ -deficiency index  $(0, p)$ . Hence, by Theorem 4.2,  $H(V)$  has deficiency index  $(0, p)$  if and only if  $\mathfrak{R}(V - F) \supseteq \mathfrak{M}_A^-$ , as we wished to prove.

On the other hand, if we make use of the resolution  $\{f, H(V)f, Af\} = \{f_+, H^*f_+, Af_+\} + \{f^+, if^+, Af^+\} + \{f_0, Hf_0, 0\}$ , where the first component is in  $\mathfrak{B}_+(A)$ , and the second in  $\mathfrak{B}^+(A)$ , also provided in Theorem 3.9, an entirely similar argument yields the result that  $H(V)$  has deficiency index  $(p, 0)$  if and only if  $\mathfrak{R}(V^{-1} - F^*) \supseteq \mathfrak{M}_A^+$ . We leave the details here to the reader.

Since  $H(V)$  is self-adjoint if and only if it has deficiency index  $(0, 0)$ , the concluding assertion of the first paragraph of the theorem follows at once.

To prove the proposition formulated in the second paragraph of the theorem, we consider an arbitrary maximal symmetric extension  $S$  of  $H$ ,  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$ , and denote by  $\mathfrak{R}$  the closure of  $A\mathfrak{B}(S)$ . Then, if  $V$  is the isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $\mathfrak{R} = \mathfrak{R}(I - V)$ ,  $V$  is closed by Theorem 2.2,  $S = H(V)$ , and the boundary condition  $Af \in \mathfrak{R}(I - V)$  is obviously nondegenerate since  $A\mathfrak{B}(S)$  has  $\mathfrak{R}(I - V)$  for its closure. Now suppose that  $S = H(V_1)$  and that  $V_1$  is closed. Then  $\mathfrak{R}(I - V_1)$  is closed and, since  $\mathfrak{R}(I - V_1) \supseteq A\mathfrak{B}(S)$ , we have  $\mathfrak{R}(I - V_1) \supseteq \mathfrak{R}(I - V)$ . Hence, by Theorem 2.2,  $V_1 \supseteq V$ . But if  $V_1 \supset V$ ,  $\mathfrak{D}(V_1)$  contains an element  $h \neq 0$  perpendicular to  $\mathfrak{D}(V)$  and  $h - V_1 h$  is thus perpendicular to  $\mathfrak{R}(I - V)$ . Furthermore there can be no element in  $\mathfrak{R}(I - V_1) \cdot \mathfrak{R}(A)$  which is not in  $\mathfrak{R}(I - V)$ , since otherwise  $S$  would not be maximal. Therefore  $V_1 \supset V$  implies that the condition  $Af \in \mathfrak{R}(I - V_1)$  is degenerate, and the proof is complete.

That when  $A$  is unbounded there do exist degenerate boundary conditions defining maximal symmetric extensions of  $H$  will be proved in the next section.

It is now easy to prove, by arguments of the same tenor as those used to establish the first portion of the preceding theorem, the following statement:

**THEOREM 4.9.** *A necessary and sufficient condition that the transformation  $H(V)$  of Definition 4.1 have a maximal symmetric closure with deficiency index  $(0, p)$  ( $(p, 0)$ ) is that the transformation  $H(\mathfrak{R}(V - F))$  have  $H(\mathfrak{M}^-)$  for its closure (that the transformation  $H(\mathfrak{R}(V^{-1} - F^*))$  have  $H(\mathfrak{M}^+)$  for its closure).*

We leave the demonstration to the reader, pausing only to point out that  $H(\mathfrak{M}^+) - iI$  and  $H(\mathfrak{M}^-) + iI$  both have bounded inverses, each with domain  $\mathfrak{S}$ , so that  $H(\mathfrak{M}^+)$  and  $H(\mathfrak{M}^-)$  are necessarily closed.

As we shall show later, it is possible to have  $\tilde{H}(V_1) \equiv \tilde{H}(V_2)$ ,  $\tilde{V}_1 \neq \tilde{V}_2$ , while  $\tilde{H}(V_1)$  is maximal. In view of this situation, it is desirable to supplement Theorem 4.9 with the following proposition:

**THEOREM 4.10.** *Let  $S$  be an arbitrary maximal symmetric extension of  $H$  such that  $\mathfrak{B}(S) \cdot \mathfrak{D}(A)$  has  $\mathfrak{B}(S)$  for its closure and let  $S_A$  be the contraction of  $S$  with graph  $\mathfrak{B}(S) \cdot \mathfrak{D}(A)$ . Then  $\tilde{S}_A \equiv S$  and there exists one and only one closed isometric transformation  $V$  with the following properties:*

- (1) *the boundary condition  $Af \in \mathfrak{R}(I - V)$  is nondegenerate;*
- (2) *for every proper isometric extension  $V_1$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  of  $V$ , the boundary condition  $Af \in \mathfrak{R}(I - V_1)$  is degenerate;*
- (3)  $S_A = H(V)$ .

The relation  $\tilde{S}_A = S$  follows at once from the fact that the closure of the graph of  $S_A$  is the graph of  $S$ .

Now let  $\mathfrak{R}$  be the closure of  $A\mathfrak{B}(S_A)$ . Then, by Theorem 2.2, there exists a closed isometric transformation  $V$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $\mathfrak{R} = \mathfrak{R}(I - V)$ . Thus the condition  $Af \in \mathfrak{R}(I - V)$  is nondegenerate and  $S_A = H(V)$ , so that (1) and (3) are satisfied.

Now suppose  $V \subset V_1$  and the condition  $Af \in \mathfrak{R}(I - V_1)$  is nondegenerate. Then there must exist an element  $h^+$  of  $\mathfrak{D}(V_1)$  which is not in  $\mathfrak{D}(V)$ , such that  $h^+ - Vh^+$  is in  $\mathfrak{R}(A)$ . Hence  $H(V_1) \supset H(V)$ , and, since  $\tilde{H}(V)$  is maximal symmetric,  $\tilde{H}(V_1) = \tilde{H}(V)$ . But then  $S = \tilde{H}(V_1)$  and, since  $S_A = H(V)$  and the graph of  $H(V_1)$  is in  $\mathfrak{D}(A)$ , we have, by definition of  $S_A$ ,  $H(V) \supseteq H(V_1)$ , which is incompatible with the inequality  $H(V_1) \supset H(V)$ . Consequently, we must conclude that the condition  $Af \in \mathfrak{R}(I - V_1)$  is degenerate; therefore (2) is satisfied by the transformation  $V$ .

To conclude the proof, we have only to note that if  $V_1$  is an arbitrary closed isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $S_A = H(V_1)$ , we have clearly  $V_1 \supseteq V$ ; hence, since we have already shown that  $V \subset V_1$  implies that the condition  $Af \in \mathfrak{R}(I - V_1)$  is degenerate, there exists only one closed isometric transformation  $V$  satisfying conditions (1)-(3).

It is to be emphasized that the equations  $\tilde{\mathfrak{R}}(V - F) = \mathfrak{M}^-$ ,  $\tilde{\mathfrak{R}}(V^{-1} - F^*) = \mathfrak{M}^+$  are necessary but not sufficient, respectively, for the conditions

$$\tilde{H}(\mathfrak{R}(V - F)) = H(\mathfrak{M}^-), \quad \tilde{H}(\mathfrak{R}(V^{-1} - F^*)) = H(\mathfrak{M}^+)$$

of Theorem 4.9. A portion of the interest of the following theorem derives from this fact.

**THEOREM 4.11.** *Let  $V$  be an isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ , and let  $F$  have the same meaning as in Theorem 3.11. Let  $K_V$  be the transformation  $(I - VF^*)^{-1}(V - F)$ . Then if  $(\mathfrak{M}^- \ominus \tilde{\mathfrak{R}}(K_V)) \cdot \mathfrak{M}_A^- = \mathfrak{D}$ ,  $H(V)$  is maximal symmetric with deficiency index  $(0, p)$  and if  $(\mathfrak{M}^+ \ominus \tilde{\mathfrak{D}}(K_V)) \cdot \mathfrak{M}_A^+ = \mathfrak{D}$ ,  $H(V)$  is maximal symmetric with deficiency index  $(p, 0)$ ; furthermore, if  $X$  is the isometric transformation from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that*

$$\mathfrak{B}(H(V)) \cdot (\mathfrak{B}^+ + \mathfrak{B}^-) = \mathfrak{R}(I - X),$$

*then  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^+$  is dense in  $\mathfrak{D}(\tilde{X})$  and  $\mathfrak{R}(X) \cdot \mathfrak{B}_A^-$  is dense in  $\mathfrak{R}(\tilde{X})$ .*

*Finally, all of the preceding statements remain true if  $K_V$  is the transformation  $(V^{-1} - F^*)^{-1}(I - V^{-1}F)$ .*

Let  $h^+$  be an arbitrary element of the domain of

$$K_V \equiv (I - VF^*)^{-1}(V - F).$$

and let  $h^- = K_V h^+$ . Then  $(V - F)h^+ = k^-$  and  $(I - VF^*)h^- = k^-$ , where  $k^-$  is some element of  $\mathfrak{M}^-$ . Thus

$$Vh^+ - Fh^+ = h^- - VF^*h^-,$$

or

$$V(h^+ + F^*h^-) = h^- + Fh^+.$$

Therefore

$$h^+ + F^*h^- - h^- - Fh^+ = (h^+ - Fh^+) - (h^- - F^*h^-)$$

is in  $\mathfrak{R}(I - V)$ . Furthermore, by the definition of  $F$  and  $F^*$  in Theorem 3.11,  $h^+ - Fh^+$  and  $h^- - F^*h^-$  are in  $\mathfrak{R}(A)$ ,  $A^{-1}(h^+ - Fh^+)$  is in  $\mathfrak{B}_A^-$ , and  $A^{-1}(h^- - F^*h^-)$  in  $\mathfrak{B}_A^+$ . Hence, if  $\{f^-, -if^-\} = A^{-1}(h^+ - Fh^+)$  and  $\{f^+, if^+\} = A^{-1}(h^- - F^*h^-)$ , then  $(h^+ - Fh^+) - (h^- - F^*h^-)$  is in  $\mathfrak{R}(A) \cdot \mathfrak{R}(I - V)$  and

$$A^{-1}[(h^+ - Fh^+) - (h^- - F^*h^-)] = \{f^-, -if^-\} - \{f^+, if^+\}$$

is in  $\mathfrak{B}(H(V))$ . Thus  $\{f^+, if^+\}$  is in  $\mathfrak{D}(X)$  and  $\{f^-, -if^-\} = X\{f^+, if^+\}$ , where  $X$  has the meaning given in the theorem. In consequence of this result, we have the relations

$$(4.6) \quad \mathfrak{D}(X) \cdot \mathfrak{B}_A^+ \supseteq A^{-1}(I - F^*)\mathfrak{R}(K_V),$$

$$(4.7) \quad \mathfrak{R}(X) \cdot \mathfrak{B}_A^- \supseteq A^{-1}(I - F)\mathfrak{D}(K_V).$$

Now let  $\{g^+, ig^+\}$  be an element of  $\mathfrak{B}^+$  such that

$$(f^+, ig^+) - (if^+, g^+) = -i(\{f^+, if^+\}, \{g^+, ig^+\}) = 0$$

for all  $\{f^+, if^+\}$  in  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^+$  and thus, in view of (4.6), for all  $\{f^+, if^+\}$  which are in  $A^{-1}(I - F^*)\mathfrak{R}(K_V)$ . Then if  $\{g^-, -ig^-\} = G\{g^+, ig^+\}$ , we have

$$(f^+, ig^+ + ig^-) - (if^+, g^+ - g^-) = 0.$$

Hence, from our fundamental formula, since  $\{g^+ - g^-, ig^+ + ig^-\}$  is in  $\mathfrak{D}(A)$ , we have

$$(A\{f^+, if^+\}, WA\{g^+ - g^-, ig^+ - ig^-\}) = 0,$$

for all  $A\{f^+, if^+\}$  in  $(I - F^*)\mathfrak{R}(K)$ . But  $WA\{g^+ - g^-, ig^+ - ig^-\}$  is in  $\mathfrak{M}_A^-$ , by Theorem 3.11, (3), and  $\mathfrak{R}(F^*)$  is in  $\mathfrak{M}^+ = \mathfrak{M} \ominus \mathfrak{M}^-$ . Thus, if

$$(\mathfrak{M}^- \ominus \overline{\mathfrak{R}(K_V)}) \cdot \mathfrak{M}_A^- = \mathfrak{D},$$

we must have

$$WA\{g^+ - g^-, ig^+ - ig^-\} = 0,$$

whence it follows at once that  $\{g^+, ig^+\} = 0$ . Hence, if

$$(\mathfrak{M}^- \ominus \overline{\mathfrak{R}(K_V)}) \cdot \mathfrak{M}_A^- = \mathfrak{D},$$

then  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^-$  has  $\mathfrak{B}^+$  for its closure. Therefore  $\mathfrak{B}^+ = \mathfrak{D}(\tilde{X})$ , and, by Theorems 2.2 and 4.2,  $H(V)$  is maximal symmetric with deficiency index  $(0, p)$ , as we wished to prove. Moreover,  $\mathfrak{R}(X) \cdot \mathfrak{B}_A^-$  is clearly  $X[\mathfrak{D}(X) \cdot \mathfrak{B}_A^+]$  and therefore has  $\mathfrak{R}(\tilde{X})$  for its closure; since we have already shown that  $\mathfrak{D}(X)$  is the closure of  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^+$ .

Thus our statements concerning the condition  $(\mathfrak{M}^- \ominus \overline{\mathfrak{R}}(K_V) \cdot \mathfrak{M}_A^- = \mathfrak{D})$  are completely established, for  $K_V \equiv (I - VF^*)^{-1}(V - F)$ . Since the condition  $(\mathfrak{M}^+ \ominus \overline{\mathfrak{D}}(K_V) \cdot \mathfrak{M}_A^+ = \mathfrak{D})$  can be discussed along entirely similar lines, making use of the relation (4.7) instead of (4.6), we leave this portion of the argument to the reader.

Now let us suppose that  $K_V \equiv (V^{-1} - F^*)^{-1}(I - V^{-1}F)$ . Again let  $h^+$  be an element of the domain of  $K_V$  and let  $h^- = K_V h^+$ . Then  $(I - V^{-1}F)h^+ = k^+$  and  $(V^{-1} - F^*)h^- = k^+$ , where  $k^+$  is some element of  $\mathfrak{M}^+$ . Then

$$h^+ - V^{-1}Fh^+ = V^{-1}h^- - F^*h^-,$$

or

$$V^{-1}(h^- + Fh^+) = h^+ + F^*h^-,$$

and

$$(h^- + Fh^+) = V(h^+ + F^*h^-).$$

Since this is an equation which occurs in the proof for the case

$$K_V \equiv (I - VF^*)^{-1}(V - F),$$

and which, together with arguments not involving  $K_V$ , leads to the relations (4.6) and (4.7), the rest of the demonstration now proceeds as before.

While we have not investigated all the questions involved, the indications are that Theorem 4.11 has no precise converse: as far as we can determine, it is possible to have  $\tilde{H}(V) \equiv \tilde{H}(V_1)$ ,  $V_1 \neq V$ , while  $V$ , but not  $V_1$ , satisfies one of the conditions of Theorem 4.10. We do, however, have the following theorem:

**THEOREM 4.12.** *Let  $S$  and  $S_A$  be as in Theorem 4.10, and let  $V$  be the isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  associated with  $S$  by that theorem. Let  $X$  be the isometric transformation from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that  $\mathfrak{R}(I - X) = \mathfrak{B}(S_A) \cdot (\mathfrak{B}^+ + \mathfrak{B}^-)$ , and let  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^+$  be dense in  $\mathfrak{D}(X)$ , or, equivalently, let  $\mathfrak{R}(X) \cdot \mathfrak{B}_A^-$  be dense in  $\mathfrak{R}(X)$ . Let  $V_1$  be an arbitrary maximal isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $V_1 \supseteq V$ , and let  $K_{V_1}$  be either*

$$(I - V_1F^*)^{-1}(V_1 - F)$$

or

$$(V^{-1} - F^*)^{-1}(I - V^{-1}F)$$

according as  $\mathfrak{D}(V_1) = \mathfrak{M}^+$  or not. Then

$$(\mathfrak{M}^- \ominus \overline{\mathfrak{R}}(K_{V_1})) \cdot \mathfrak{M}_A^- = \mathfrak{D}$$

if the deficiency index of  $S$  is  $(0, p)$ , and

$$(\mathfrak{M}^+ \ominus \overline{\mathfrak{D}}(K_{V_1})) \cdot \mathfrak{M}_A^+ = \mathfrak{D}$$

if the deficiency index of  $S$  is  $(p, 0)$ .

Let  $\{f^+, if^+\}$  be an arbitrary element of  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^+$ , and let

$$\{f^-, -if^-\} = X\{f^+, if^+\}.$$

Then  $\{f^+ - f^-, if^+ + if^-\}$  is in  $\mathfrak{D}(A)$  and

$$\begin{aligned} A\{f^+ - f^-, if^+ + if^-\} &= A\{f^+, if^+\} - A\{f^-, -if^-\} \\ &= -h^- + F^*h^- + h^+ - Fh^+, \end{aligned}$$

where  $h^-$  is in  $\mathfrak{M}^-$ ,  $h^+$  in  $\mathfrak{M}^+$ . But  $\{f^+ - f^-, if^+ + if^-\}$  is in  $\mathfrak{B}(S)$  also; so we have

$$h^- + Fh^+ = V_1(F^*h^- + h^+).$$

Thus, if  $\mathfrak{D}(V_1) = \mathfrak{M}^+$ ,

$$h^- + Fh^+ = V_1F^*h^- + V_1h^+,$$

or, solving for  $h^-$ ,

$$h^- = (I - V_1F^*)^{-1}(V_1 - F)h^+ = K_{V_1}h^+.$$

On the other hand, if  $\mathfrak{D}(V_1) \neq \mathfrak{M}^+$ , we must have  $\mathfrak{R}(V_1) = \mathfrak{M}^-$ , so that

$$V_1^{-1}h^- + V_1^{-1}Fh^+ = F^*h^- + h^+.$$

Thus, solving for  $h^-$ , we have in this case

$$h^- = (V_1^{-1} - F^*)^{-1}(I - V_1^{-1}F)h^+ = K_{V_1}h^+.$$

Therefore, in both cases,  $h^- = K_{V_1}h^+$ .

Now let  $k^-$  be an element of  $\mathfrak{M}_A^-$  which is perpendicular to  $\overline{\mathfrak{R}}(K_{V_1})$ . Then, if  $A_1^{-1}k^- = \{g^+ - g^-, ig^+ + ig^-\}$ , we have

$$(f^+, ig^+ + ig^-) - (if^+, g^+ - g^-) = -(h^- - F^*h^-, Wk^-) = -i(h^- - F^*h^-, k^-),$$

where  $f^+$  and  $h^-$  have the same meanings as above; and this equation is equivalent to

$$(f^+, ig^+) - (if^+, g^+) = -i(h^-, k^-),$$

since  $F^*h^-$  is in  $\mathfrak{M}^+$ . Hence, since  $k^-$  is perpendicular to  $\overline{\mathfrak{R}}(K_{V_1})$ ,

$$(\{f^+, if^+\}, \{g^+, ig^+\}) = 0.$$

Now suppose that  $S$  has deficiency index  $(0, p)$ . Then  $\mathfrak{D}(\tilde{X}) = \mathfrak{B}^+$ . Hence, if  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^+$  is dense in  $\mathfrak{D}(\tilde{X})$ , we must have  $\{g^+, ig^+\} = 0$ , since  $\{f^+, if^+\}$  is an arbitrary element in  $\mathfrak{D}(X) \cdot \mathfrak{B}_A^+$ . Therefore  $k^- = 0$  and  $(\mathfrak{M}^- \ominus \overline{\mathfrak{R}(K_{V_1})}) \cdot \mathfrak{M}_A^- = \mathfrak{D}$ , as we wished to prove.

On the other hand, if  $S$  has deficiency index  $(p, 0)$ , then  $\mathfrak{R}(\tilde{X}) = \mathfrak{B}^-$ ; and an entirely similar argument leads to the conclusion  $(\mathfrak{M}^+ \ominus \overline{\mathfrak{D}(K_{V_1})}) \cdot \mathfrak{M}_A^+ = \mathfrak{D}$ .

Finally, we note that if  $\mathfrak{R}(I - X) \subseteq \mathfrak{D}(A)$ , then  $X\{f^+, if^+\}$  is in  $\mathfrak{B}_A^-$  whenever  $\{f^+, if^+\}$  is in  $\mathfrak{B}_A^+$ . Hence  $\mathfrak{R}(\tilde{X})$  is clearly the closure of  $\mathfrak{R}(X) \cdot \mathfrak{B}_A^-$  whenever  $\mathfrak{D}(\tilde{X})$  is the closure of  $\mathfrak{D}(\tilde{X}) \cdot \mathfrak{B}_A^+$ , as indicated in the theorem.

All of the Theorems 4.8–4.12 apply, of course, to the case in which  $A$  is bounded as well as to the case in which  $A$  is unbounded. For the former case, however, they are superficial, in view of Theorem 4.1, whereas in the latter they are not. The reason for this is to be found in Theorem 3.11, which tells us that the transformation  $F$  has a bound less than unity when  $A$  is bounded, and the bound unity which it never attains when  $A$  is unbounded. Consequently, if  $A$  is bounded and  $V$  is an arbitrary isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ ,  $V - F$  and  $V^{-1} - F^*$  have bounded inverses. On the other hand, if  $A$  is unbounded,  $(V - F)^{-1}$  and  $(V^{-1} - F^*)^{-1}$  always exist, but are in general unbounded.

**5. Pathology of unbounded reduction operators.** We now examine certain pathological aspects of the behavior of discontinuous reduction operators, with a view to revealing more clearly the significance of the results of the preceding section.

To begin, we state and prove three lemmas, the roles of which will become clear later.

**LEMMA 4.2.** *Let  $\mathfrak{D}$  be the domain, everywhere dense in a Hilbert space  $\mathfrak{H}$ , of a closed linear unbounded transformation  $R$  with range in a Hilbert space  $\mathfrak{H}_0$ . Then there exists a Hilbert space  $\mathfrak{U}$  in  $\mathfrak{H}$  such that  $\mathfrak{U} \cdot \mathfrak{D} = \mathfrak{D}$ .*

*If  $R$  has range  $\mathfrak{H}_0$  and  $R^{-1}$  exists, then the set of elements  $f$  of  $\mathfrak{D}(R^*)$  for which  $R^*f$  is in  $\mathfrak{H} \ominus \mathfrak{U}$ , is dense in  $\mathfrak{H}_0$ .*

According to a theorem previously noted, we can determine a self-adjoint transformation  $T$  in  $\mathfrak{H}$ , with domain  $\mathfrak{D}$ .† Furthermore, there must exist two Hilbert spaces  $\mathfrak{H}_1$  and  $\mathfrak{H}_2$ ,  $\mathfrak{H}_1 + \mathfrak{H}_2 = \mathfrak{H}$ , both of which reduce  $T$ . Let  $T_k$  be the transformation induced by  $T$  in  $\mathfrak{H}_k$ , ( $k = 1, 2$ ). Then at least one of the transformations  $T_1, T_2$ , say  $T_1$ , is unbounded. Let  $T_3$  be the linear transformation in  $\mathfrak{H}$  which is equal to  $T_1$  in  $\mathfrak{H}_1$ , to  $I$  in  $\mathfrak{H}_2$ . Then  $T_3$  is clearly unbounded self-adjoint and  $\mathfrak{D}(T_3) \supseteq \mathfrak{D}(T)$ ,  $\mathfrak{D}(T_3) \supseteq \mathfrak{H}_2$ . Furthermore, there exists a unitary transformation  $U$  in  $\mathfrak{H}$  such that  $U\mathfrak{D}(T_3)$  and  $\mathfrak{D}(T_3)$  intersect only in the

† Murray, Theorem 1.24.

origin.† Hence the Hilbert space  $\mathfrak{U} = U\mathfrak{E}_2$  intersects the manifold  $\mathfrak{D}$  only in the origin.

To prove the final assertion of the lemma, we consider an arbitrary element  $g$  of  $\mathfrak{E}_0$  which satisfies the equation  $(g, f) = 0$  for all  $f$  in  $\mathfrak{D}(R^*)$  such that  $R^*f$  is in  $\mathfrak{E} \ominus \mathfrak{U}$ ; if  $R$  has range  $\mathfrak{E}_0$  and  $R^{-1}$  exists, then  $R^{-1}g$  is defined and  $(R^{-1}g, R^*f) = 0$  for all  $R^*f$  in  $\mathfrak{E} \ominus \mathfrak{U}$ . But if  $R^{-1}$  exists and has domain  $\mathfrak{E}_0$ , it follows also that  $R^*$  has range  $\mathfrak{E}$ , and thus  $\mathfrak{R}(R^*) \supset \mathfrak{E} \ominus \mathfrak{U}$ . Hence  $R^{-1}g$  is in  $\mathfrak{D} \cdot \mathfrak{U}$  and thus  $R^{-1}g = 0$ . Consequently  $g = 0$ , and the proof is complete.

**LEMMA 4.3.** *Let  $Q$  be a unitary transformation in a Hilbert space  $\mathfrak{E}$  such that  $Q^2 + I \equiv 0$ ; and let the characteristic manifolds  $\mathfrak{E}^+$  and  $\mathfrak{E}^-$  of  $Q$ , for the characteristic values  $+i$  and  $-i$ , respectively, both be Hilbert spaces. Let  $D$  be an unbounded nonnegative definite self-adjoint transformation in  $\mathfrak{E}$  such that  $D^{-1}$  exists and  $D \equiv QD^{-1}Q^{-1}$ . Then there exists a maximal  $Q$ -symmetric manifold  $\mathfrak{R}$  in  $\mathfrak{E}$ , with the following properties:*

- (1)  $\mathfrak{R}$  is in  $\mathfrak{D}(D)$  and  $D\mathfrak{R}$  has a maximal  $Q$ -symmetric closure;
- (2) for each cardinal number  $m$  on the range  $1 \leq m \leq \aleph_0$ , there exists in  $\mathfrak{R}$  a closed linear  $Q$ -symmetric manifold  $\mathfrak{R}_1$  which has the  $Q$ -deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $(0, p)$  or  $(p, 0)$  is the  $Q$ -deficiency index of  $\mathfrak{R}$ , while  $D\mathfrak{R}_1$  has the same closure as  $D\mathfrak{R}$ .

Let  $\mathfrak{U}$  be the range of  $E(1-0)$ , where  $E(\lambda)$  is the resolution of the identity in  $\mathfrak{E}$  associated with  $D$ , and let  $D_1$  be the transformation induced in  $\mathfrak{U}$  by  $D$ . Then  $D_1^{-1}$  exists,  $\mathfrak{U}$  is  $Q$ -symmetric, and in  $Q\mathfrak{U}$ ,  $D = QD_1^{-1}Q_1$ , by Theorem 3.5. By the same theorem,  $D$  is equal to  $I$  in  $\mathfrak{E} \ominus (\mathfrak{U} + Q\mathfrak{U})$ , and hence  $D_1^{-1}$  is unbounded since  $D$  is. Therefore  $\mathfrak{D}(D_1^{-1})$  is dense in  $\mathfrak{U}$ , but not identically  $\mathfrak{U}$ . Applying Lemma 4.2, we determine a Hilbert space  $\mathfrak{U}_0$  in  $\mathfrak{U}$  such that  $\mathfrak{D}(D_1^{-1}) \cdot \mathfrak{U}_0 = \mathfrak{D}$ . Then, by the second paragraph of the same lemma,  $D(\mathfrak{U} \ominus \mathfrak{U}_0)$  is dense in  $\mathfrak{U}$ . Hence, if  $\mathfrak{U}_1$  is any closed linear subspace of  $\mathfrak{U}_0$ , with dimension number  $m$ ,  $D(\mathfrak{U} \ominus \mathfrak{U}_1)$  is dense in  $\mathfrak{U}$ . Now let  $\mathfrak{R}$  be an arbitrary maximal  $Q$ -symmetric extension of  $\mathfrak{U}$ , and let  $\mathfrak{R}_1 = (\mathfrak{U} \ominus \mathfrak{U}_1) + (\mathfrak{R} \ominus \mathfrak{U})$ . Then, since  $\mathfrak{R} \ominus \mathfrak{U}$  is in the manifold  $\mathfrak{E} \ominus (\mathfrak{R} + Q\mathfrak{R})$ , where  $D = I$ , it follows that  $D\mathfrak{R}$  and  $D\mathfrak{R}_1$ , both have the closure  $\mathfrak{R}$ . As  $\mathfrak{R}_1$  evidently has the  $Q$ -deficiency index stated in the lemma, the proof is complete.

**LEMMA 4.4.** *Let  $\mathfrak{E}$ ,  $\mathfrak{E}^+$ ,  $\mathfrak{E}^-$ ,  $Q$ , and  $D$  be as in Lemma 4.3. Let  $m$  be an arbitrary cardinal number on the range  $0 \leq m \leq \aleph_0$ . Then there exists a maximal  $Q$ -symmetric manifold  $\mathfrak{R}$  in  $\mathfrak{D}(D)$  such that  $D\mathfrak{R}$  is a closed linear  $Q$ -symmetric manifold with the  $Q$ -deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $(0, p)$  or  $(p, 0)$  is the  $Q$ -deficiency index of  $\mathfrak{R}$ .*

† Von Neumann, Journal für die reine und angewandte Mathematik (Crelle), loc. cit.

Let  $\mathfrak{U}$  and  $D_1$  have the same meanings as in the proof of Lemma 4.3. Let  $V$  be the isometric transformation with domain in  $\mathfrak{E}^+$  and range in  $\mathfrak{E}^-$ , such that  $\mathfrak{U} = \mathfrak{R}(I - V)$ . Then it is readily shown that  $D_1$  determines a unique bounded self-adjoint transformation  $D_0$  in  $\mathfrak{D}(V)$ , such that for each  $h^+ - Vh^+$  in  $\mathfrak{U}$ ,

$$D(h^+ - Vh^+) = D_0h^+ - VD_0h^+.$$

Moreover,  $D_0$  has the same bound, 1, as  $D_1$ , and  $D_0^{-1}$  exists and is unbounded because  $D_1^{-1}$  is unbounded.

Therefore, applying Lemma 4.2, we can determine a bounded closed contraction  $C$  of  $D_0$  such that  $\mathfrak{D}(V) \ominus \mathfrak{D}(C)$  has the dimension number  $m$ , ( $0 \leq m \leq \aleph_0$ ), and such that  $\mathfrak{R}(C)$  is dense in  $\mathfrak{D}(V)$ . Hence  $C^{-1}$ , which exists because  $D_0^{-1}$  does, is a closed symmetric transformation in  $\mathfrak{D}(V)$ , and  $R \equiv [(C^{-1})^*C^{-1}]^{1/2}$  is a self-adjoint transformation with the same range as  $(C^{-1})^*$ . Since  $D_0^{-1} \supseteq C^{-1}$  and  $D_0^{-1} \equiv (D_0^{-1})^*$ , it follows that  $D_0^{-1} \subseteq (C^{-1})^*$  and hence that  $(C^{-1})^*$  has range identically  $\mathfrak{D}(V)$ . Thus  $R$  has range  $\mathfrak{D}(V)$  and domain  $\mathfrak{R}(C)$ .<sup>†</sup> Consequently as  $R$  is self-adjoint,  $R^{-1}$  exists.

Now consider the set  $\mathfrak{N}_1$  of all elements of  $\mathfrak{E}$  which can be written in the form

$$(I + iR^{-1})h^+ - V(I - iR^{-1})h^+.$$

Since  $I + iR^{-1} \equiv i(R^{-1} - iI)$ , and  $R^{-1}$  is self-adjoint in  $\mathfrak{D}(V)$ ,  $(I + iR^{-1})$  has range identically  $\mathfrak{D}(V)$  and  $\mathfrak{N}_1$  is precisely the set of all elements of  $\mathfrak{E}$  which can be written in the form

$$k^+ - V(I - iR^{-1})(I + iR^{-1})^{-1}k^+,$$

where  $k^+$  is in  $\mathfrak{D}(V)$ . But  $(I - iR^{-1})(I + iR^{-1})^{-1}$  is evidently a unitary transformation in  $\mathfrak{D}(V)$  and thus  $V(I - iR^{-1})(I + iR^{-1})^{-1}$  is an isometric transformation with domain  $\mathfrak{D}(V)$  and range  $\mathfrak{R}(V)$ . Hence  $\mathfrak{N}_1$  is closed  $Q$ -symmetric, by Theorem 2.2.

Next, let  $\mathfrak{N}$  be any maximal  $Q$ -symmetric extension of  $\mathfrak{N}_1$ . Then  $\mathfrak{N} \ominus \mathfrak{N}_1$  belongs to the manifold where  $D = I$  and thus is in  $\mathfrak{D}(D)$ . Furthermore, every element

$$(I + iR^{-1})h^+ - V(I - iR^{-1})h^+$$

of  $\mathfrak{N}_1$  can be expressed in the form

$$(h^+ - Vh^+) + i(R^{-1}h^+ + VR^{-1}h^+).$$

But  $R^{-1}h^+$  is in the range of  $D_0$  by definition of  $R^{-1}$  and  $R^{-1}h^+ - VR^{-1}h^+$  is

<sup>†</sup> Murray, Theorem 1.24.

thus in the range of  $D_1$ . Hence, since  $\mathfrak{R}(I+V) = Q\mathfrak{R}$  and since, in  $Q\mathfrak{R}$ ,  $D = QD_1^{-1}Q^{-1}$ , it follows that  $R^{-1}h^+ + VR^{-1}h^+$  is in  $\mathfrak{D}(D)$ . Hence  $\mathfrak{R}$  is in  $\mathfrak{D}(D)$ . Moreover,  $D\mathfrak{R}$  and  $D\mathfrak{R}_1$  are easily shown to be linear  $Q$ -symmetric.

Now, to determine the  $Q$ -deficiency index of  $D\mathfrak{R}_1$ , we again take account of the fact that  $\mathfrak{R}_1$  is in  $\mathfrak{U} + Q\mathfrak{U} = \mathfrak{D}(V) + \mathfrak{R}(V)$ . Since this manifold reduces  $D$ ,  $D\mathfrak{R}_1$  also belongs to it; moreover,  $\mathfrak{R} \ominus \mathfrak{R}_1$  reduces  $D$  since it belongs to  $\mathfrak{R} \ominus (\mathfrak{D}(V) + \mathfrak{R}(V))$ . Consequently, there exists a unique isometric transformation  $X$  with domain in  $\mathfrak{D}(V)$  and range in  $\mathfrak{R}(V)$  such that  $D\mathfrak{R}_1 = \mathfrak{R}(I - X)$ ; and the  $Q$ -deficiency index of  $D\mathfrak{R}_1$  is precisely  $(n, q + p)$  or  $(n + p, q)$ , where  $n$  and  $q$  are the dimension numbers of  $\mathfrak{D}(V) \ominus \mathfrak{D}(X)$  and  $\mathfrak{R}(V) \ominus \mathfrak{R}(X)$ , respectively, according as  $(0, p)$  or  $(p, 0)$  is the  $Q$ -deficiency index of  $\mathfrak{R}$ . Furthermore,  $D\mathfrak{R}$  is closed if and only if  $D\mathfrak{R}_1$  is closed and  $D\mathfrak{R}_1$  is closed if and only if  $X$  is closed.

Thus to complete the proof it remains only to be shown that  $X$  is closed and that  $\mathfrak{D}(V) \ominus \mathfrak{D}(X)$  and  $\mathfrak{R}(V) \ominus \mathfrak{R}(X)$  both have the dimension number  $m$ .

To determine  $X$ , we begin by analyzing further the manifold  $D\mathfrak{R}_1$ , using the resolution

$$h^+ - Vh^+ + i(R^{-1}h^+ + VR^{-1}h^+)$$

for an element of  $\mathfrak{R}_1$ . Since

$$D(h^+ - Vh^+) = D_0h^+ - VD_0h^+$$

and

$$iD(R^{-1}h^+ + VR^{-1}h^+) = i(D_0^{-1}R^{-1}h^+ + VD_0^{-1}R^{-1}h^+),$$

every element of  $D\mathfrak{R}_1$  can be written in the form

$$(D_0 + iD_0^{-1}R^{-1})h^+ - V(D_0 - iD_0^{-1}R^{-1})h^+.$$

Conversely, every element of this form is easily shown to be in  $D\mathfrak{R}_1$ . Thus, since  $(D_0 + iD_0^{-1}R^{-1})h^+$  is in  $\mathfrak{D}(V)$  and  $V(D_0 - iD_0^{-1}R^{-1})h^+$  is in  $\mathfrak{R}(V)$ , it follows that

$$\mathfrak{D}(X) = \mathfrak{R}(D_0 + D_0^{-1}R^{-1})$$

and that

$$X(D_0 + iD_0^{-1}R^{-1})h^+ = V(D_0 - iD_0^{-1}R^{-1})h^+.$$

Hence, as  $(D_0 + iD_0^{-1}R^{-1})^{-1}$  is readily shown to exist,

$$X \equiv V(D_0 - iD_0^{-1}R^{-1})(D_0 + iD_0^{-1}R^{-1})^{-1},$$

or, by straightforward algebraic calculation,

$$X \equiv -V(D_0^{-1}R^{-1}D_0^{-1} + iI)(D_0^{-1}R^{-1}D_0^{-1} - iI)^{-1},$$

and

$$-V^{-1}X \equiv (D_0^{-1}R^{-1}D_0^{-1} + iI)(D_0^{-1}R^{-1}D_0^{-1} - iI)^{-1}.$$

We now recall that  $D_0^{-1}$  has range  $\mathfrak{D}(V)$  and that  $R^{-1}$  has domain  $\mathfrak{D}(V)$ , while  $\mathfrak{D}(R) = \mathfrak{R}(R^{-1})$  is in  $\mathfrak{D}(D_0^{-1})$ . Thus, it is evident that  $D_0^{-1}R^{-1}D_0^{-1}$  is symmetric because its domain is dense in  $\mathfrak{D}(V)$ , and  $D_0^{-1}$  and  $R^{-1}$  are both self-adjoint. Furthermore,  $-V^{-1}X$  is the Cayley transform of  $D_0^{-1}R^{-1}D_0^{-1}$  from the equation at the end of the preceding paragraph. Therefore,  $X$  is closed if and only if  $D_0^{-1}R^{-1}D_0^{-1}$  is closed and, if  $X$  is closed, the dimension numbers of  $\mathfrak{D}(V) \ominus \mathfrak{D}(X)$  and  $\mathfrak{R}(V) \ominus \mathfrak{R}(X)$  are  $n$  and  $q$ , respectively, where  $(n, q)$  is the deficiency index of  $D_0^{-1}R^{-1}D_0^{-1}$ . Consequently, we can complete the proof by showing that  $D_0^{-1}R^{-1}D_0^{-1}$  is closed and that its deficiency index is  $(m, m)$ .

To prove the first we begin by observing that

$$\mathfrak{R}(D_0^{-1}R^{-1}D_0^{-1}) = D_0^{-1}\mathfrak{D}(R)$$

is a closed linear manifold by definition of  $R$ . We next observe that  $D_0^{-1}$  and  $R^{-1}$  are positive definite, each with lower bound 1; in consequence, we have

$$(D_0^{-1}R^{-1}D_0^{-1}h^+, h^+) = (R^{-1}D_0^{-1}h^+, D_0^{-1}h^+) \geq (D_0^{-1}h^+, D_0^{-1}h^+) \geq |h^+|^2.$$

Thus  $D_0^{-1}R^{-1}D_0^{-1}$  is positive definite with lower bound greater than or equal to 1. Hence it is easily shown that  $(D_0^{-1}R^{-1}D_0^{-1})^{-1}$  exists and is bounded. But, as we have already pointed out,  $D_0^{-1}R^{-1}D_0^{-1}$  has a closed range; hence its inverse is closed and in consequence  $D_0^{-1}R^{-1}D_0^{-1}$  itself is closed.

We have now only to show that the deficiency index of  $D_0^{-1}R^{-1}D_0^{-1}$  is  $(m, m)$ . Let us suppose that the deficiency index of  $D_0^{-1}R^{-1}D_0^{-1}$  is  $(n, n)$ ; that it is of this form follows from the fact that  $D_0^{-1}R^{-1}D_0^{-1}$  is positive definite. Also from this fact, it follows that  $D_0^{-1}R^{-1}D_0^{-1}$  has a self-adjoint extension  $T$  with bounded inverse; indeed,  $T$  may be chosen with the same lower bound, 1, as  $D_0^{-1}R^{-1}D_0^{-1}$ . Furthermore, from Theorems 4.2, 2.8, and 2.2, it is readily deduced that  $n$  is the dimension number of

$$\mathfrak{B}(T) \ominus \mathfrak{B}(D_0^{-1}R^{-1}D_0^{-1}).$$

But from the fact that  $T^{-1}$  and  $(D_0^{-1}R^{-1}D_0^{-1})^{-1}$  both exist and are bounded, it follows by a straightforward argument that

$$\mathfrak{B}(T) \ominus \mathfrak{B}(D_0^{-1}R^{-1}D_0^{-1})$$

and

$$\mathfrak{R}(T) \ominus \mathfrak{R}(D_0^{-1}R^{-1}D_0^{-1})$$

have the same dimension number. Moreover,  $\mathfrak{R}(D_0^{-1}R^{-1}D_0^{-1}) = \mathfrak{D}(C)$  by definition of  $\mathfrak{R}$ , and  $\mathfrak{R}(T) = \mathfrak{D}(V)$ , since  $T$  is self-adjoint in  $\mathfrak{D}(V)$ , with  $T^{-1}$

bounded. Thus, since  $\mathfrak{D}(V) \ominus \mathfrak{D}(C)$  has the dimension number  $m$  by choice of  $C$ ,  $n = m$  and the proof is complete.

**THEOREM 4.13.** *Let  $A$  be an unbounded reduction operator for  $H^*$ . Then there exists a maximal isometric transformation  $V$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  with the following properties: (1) the boundary condition  $Af \in \mathfrak{R}(I - V)$  is nondegenerate and  $H(V)$  is maximal symmetric; (2) for an arbitrary cardinal number on the range  $1 \leq m \leq \aleph_0$ , there exists an isometric transformation  $V_1$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $\tilde{V}_1 \equiv V$  and such that  $H(V_1)$  has the deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $(0, p)$  or  $(p, 0)$  is the deficiency index of  $H(V)$ .*

Let  $B \equiv (A_1^* A_1)^{1/2}$ . Then  $A_1 \equiv XB$  where  $X$  is isometric with domain  $\mathfrak{B}^+ + \mathfrak{B}^-$  and range  $\mathfrak{M}$ .† Applying Lemma 4.3, and the fact, stated in Theorem 3.6, that  $B \equiv QB^{-1}Q^{-1}$ , we conclude that there exists a maximal  $Q$ -symmetric manifold  $\mathfrak{N}$  in  $\mathfrak{B}^+ + \mathfrak{B}^-$  with the following properties:  $\mathfrak{N} \subset \mathfrak{D}(B) = \mathfrak{D}(A)$  and  $B\mathfrak{N}$  has a maximal  $Q$ -symmetric closure;  $\mathfrak{N} \supset \mathfrak{N}_1$ , where  $\mathfrak{N}_1$  is closed linear  $Q$ -symmetric with  $Q$ -deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $(0, p)$  or  $(p, 0)$  is the  $Q$ -deficiency index of  $\mathfrak{N}$ , and  $B\mathfrak{N}_1$  has the same closure as  $B\mathfrak{N}$ .

Furthermore, since, by Theorem 3.3,  $X\mathfrak{B}^+ = \mathfrak{M}^-$  and  $X\mathfrak{B}^- = \mathfrak{M}^+$ , it is clear that  $XB\mathfrak{N} = A\mathfrak{N}$  has a maximal  $W$ -symmetric closure and that  $XB\mathfrak{N}_1 = A\mathfrak{N}_1$  has the same closure as  $A\mathfrak{N}$ . Now let  $V_2$  and  $V_1$  be the isometric transformations from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  which correspond to  $A\mathfrak{N}$  and  $A\mathfrak{N}_1$ , respectively, in accordance with Theorem 2.2, and let  $V \equiv \tilde{V}_2$ . Then  $V$  is maximal isometric from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  and  $\tilde{V}_1 \equiv V$ , since  $(A\mathfrak{N}_1) = \mathfrak{N}(I - V)$ . Moreover,  $H(V)$  and  $H(V_1)$  have for their graphs  $\mathfrak{B} + \mathfrak{N}$  and  $\mathfrak{B} + \mathfrak{N}_1$ , respectively. Hence, by Theorem 4.2,  $H(V)$  is maximal symmetric and  $H(V_1)$  has the deficiency index stated in the theorem. Thus, since the boundary conditions  $Af \in \mathfrak{R}(I - V)$ ,  $Af \in \mathfrak{R}(I - V_1)$  are obviously nondegenerate, the proof is complete.

**THEOREM 4.14.** *Let  $A$  be unbounded. Then there exists a maximal isometric transformation  $V$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  with the following properties: (1) the boundary condition  $Af \in \mathfrak{R}(I - V)$  is nondegenerate and  $H(V)$  has a maximal symmetric closure; (2) if  $m$  is an arbitrary cardinal number on the range  $1 \leq m \leq \aleph_0$  and  $V$  has the  $W$ -deficiency index  $(0, p)$   $((p, 0))$ , there exists an isometric transformation  $V_1$ ,  $V_1 \subset V$ , with  $W$ -deficiency index  $(m, m+p)$   $((m+p, m))$  such that the boundary condition  $Af \in \mathfrak{R}(I - V_1)$  is nondegenerate and such that  $\tilde{H}(V_1) \equiv \tilde{H}(V)$ .*

Let  $B$  and  $X$  have the same meanings as in the proof of Theorem 4.13. Then  $B^{-1} \equiv QBQ^{-1}$ . Let  $\mathfrak{N}$  be a manifold in  $\mathfrak{D}(B^{-1})$  which satisfies the conditions of Lemma 4.3 with  $D \equiv B^{-1}$ . Then  $X\mathfrak{N}$  is clearly maximal  $W$ -symmetric

† Murray, Theorem 1.24.

in  $\mathfrak{M}$ , and if  $\mathfrak{N}_1$  has the same meaning as in Lemma 4.3, (2), then  $X\mathfrak{N}_1$  has the  $W$ -deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $X\mathfrak{N}$  has the  $W$ -deficiency index  $(0, p)$  or  $(p, 0)$ . Furthermore,  $\mathfrak{N}$  and  $\mathfrak{N}_1$  both belong to the range of  $B$ ; hence  $X\mathfrak{N}$  and  $X\mathfrak{N}_1$  belong to the range of  $A$ . Therefore the boundary conditions  $Af \in X\mathfrak{N}_1$ ,  $Af \in X\mathfrak{N}$  are nondegenerate.

Now let  $V$  and  $V_1$  be the isometric transformations from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  corresponding to  $X\mathfrak{N}$  and  $X\mathfrak{N}_1$  in accordance with Theorem 2.2. Then  $H(V)$  and  $H(V_1)$  have for their graphs  $\mathfrak{B} + B^{-1}\mathfrak{N}$  and  $\mathfrak{B} + B^{-1}\mathfrak{N}_1$ , respectively. Furthermore, by Lemma 4.3,  $B^{-1}\mathfrak{N}$  and  $B^{-1}\mathfrak{N}_1$  have the same closure and the latter is maximal  $Q$ -symmetric in  $\mathfrak{B}^+ + \mathfrak{B}^-$ . Hence, by Theorem 4.2,  $H(V)$  is maximal symmetric. Thus, since  $X\mathfrak{N}$  is maximal  $W$ -symmetric,  $V$  is maximal isometric from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ , and  $V$  and  $V_1$  have the properties described in Theorem 4.14.

In view of Theorem 4.14, the role of the condition (2) of Theorem 4.10 is now more clearly indicated.

**THEOREM 4.15.** *Let  $A$  be an unbounded reduction operator for  $H^*$ . Let  $m$  be an arbitrary cardinal number on the range  $0 \leq m \leq \aleph_0$ . Then there exists a maximal isometric transformation  $V$  from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  with the following properties:*

- (1) *the boundary condition  $Af \in \mathfrak{N}(I - V)$  is nondegenerate;*
- (2)  *$H(V)$  is closed with the deficiency index  $(m+p, m)$  or  $(m, m+p)$  according as  $(0, p)$  or  $(p, 0)$  is the  $W$ -deficiency index of  $V$ ;*
- (3) *for every maximal symmetric extension  $S$  of  $H(V)$ ,  $S_A \equiv H(V)$ , where  $S_A$  has the same meaning as in Theorem 4.5.*

Again, let  $B$  have the same meaning as in the proof of Theorem 4.13; as we have already noted,  $B^{-1} \equiv QBQ^{-1}$ . Hence, by Lemma 4.4, there exists a maximal  $Q$ -symmetric manifold  $\mathfrak{N}$  in  $\mathfrak{D}(B^{-1})$  such that  $B^{-1}\mathfrak{N}$  is closed linear  $Q$ -symmetric with  $Q$ -deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $\mathfrak{N}$  has  $Q$ -deficiency index  $(0, p)$  or  $(p, 0)$ . Furthermore, if  $A \equiv XB$ , the relations  $\mathfrak{M}^+ = X\mathfrak{B}^-$ ,  $\mathfrak{M}^- = X\mathfrak{B}^+$  imply that  $X\mathfrak{N} = A(B^{-1}\mathfrak{N})$  has the  $W$ -deficiency index  $(p, 0)$  or  $(0, p)$  according as  $\mathfrak{N}$  has the  $Q$ -deficiency index  $(0, p)$  or  $(p, 0)$ .

Now let  $V$  be isometric from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $X\mathfrak{N} = \mathfrak{N}(I - V)$ . Then, by definition, the  $W$ -deficiency index of  $V$  is the same as that of  $X\mathfrak{N}$ . Moreover, the graph of  $H(V)$  is precisely  $\mathfrak{B} + A_1^{-1}X\mathfrak{N} = \mathfrak{B} + B^{-1}\mathfrak{N}$ . Thus, by Theorem 4.2,  $H(V)$  is closed with deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $(p, 0)$  or  $(0, p)$  is the  $W$ -deficiency index of  $V$ , and the boundary condition  $Af \in \mathfrak{N}(I - V)$  is obviously nondegenerate.

We are therefore left to prove only that  $H(V)$  has the property (3). To do this, we consider an arbitrary maximal symmetric extension  $S$  of  $H(V)$  and

the transformation  $S_A$  described in Theorem 4.5. Then  $A\mathfrak{B}(S_A)$  is obviously a linear  $W$ -symmetric manifold in  $\mathfrak{M}$ ; moreover,  $A\mathfrak{B}(S_A) \supseteq A\mathfrak{B}(H(V)) = \mathfrak{R}(I-V) = \mathfrak{R}$ . But  $\mathfrak{R}$  is maximal  $W$ -symmetric; therefore we must have  $A\mathfrak{B}(S_A) = \mathfrak{R}$ , which is possible if and only if  $S_A = H(V)$ . Thus the proof is complete.

We thus see that the condition that  $V$  be maximal isometric from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  is not sufficient either for  $H(V)$  or for  $\tilde{H}(V)$  to be maximal symmetric, even when the additional restriction that the condition  $Af \in \mathfrak{R}(I-V)$  be nondegenerate is imposed. It is therefore natural to ask what properties  $H(V)$  has under these circumstances. Without giving the proofs, we shall merely set down the following observations which the reader may verify: If  $V$  is maximal isometric from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  with  $W$ -deficiency index  $(0, p)$  ( $(p, 0)$ ), if  $\mathfrak{R}(I-V)$  belongs to  $\mathfrak{R}(A)$ , and if  $X$  denotes the isometric transformation from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that

$$\mathfrak{R}(I-X) = (\mathfrak{B}^+ + \mathfrak{B}^-) \cdot \mathfrak{B}(H(V));$$

then  $\mathfrak{R}(X-G) \supseteq \mathfrak{B}_A^-$  ( $\mathfrak{R}(X^{-1}-G^*) \supseteq \mathfrak{B}_A^+$ ),  $G$  having here the same meaning as in Theorem 3.11; conversely, if  $X$  is a closed isometric transformation from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that  $\mathfrak{D}(A) \cdot \mathfrak{R}(I-X)$  has  $\mathfrak{R}(I-X)$  for its closure and if  $\mathfrak{R}(X-G) \supseteq \mathfrak{B}_A^-$  ( $\mathfrak{R}(X^{-1}-G^*) \supseteq \mathfrak{B}_A^+$ ), then  $A[\mathfrak{D}(A) \cdot \mathfrak{R}(I-X)]$  is a maximal  $W$ -symmetric manifold  $\mathfrak{R}$  in  $\mathfrak{M}$  with  $W$ -deficiency index  $(0, p)$  ( $(p, 0)$ ). The reader will note that this proposition constitutes a slightly modified analogue of Theorem 4.8. In precisely the same sense, one can formulate valid analogues of Theorems 4.9-4.12.

We shall now show that the condition that  $V$  be maximal isometric from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  and that  $Af \in \mathfrak{R}(I-V)$  be nondegenerate, is unnecessary as well as insufficient for either  $H(V)$  or  $\tilde{H}(V)$  to be maximal symmetric.

**THEOREM 4.16.** *Let  $A$  be an unbounded reduction operator for  $H^*$ . Let  $m$  be an arbitrary cardinal number on the range  $0 \leq m \leq \aleph_0$ . Then there exists a maximal symmetric extension  $S$  of  $H$  with the following properties:*

- (1)  $\mathfrak{B}(S) \subset \mathfrak{D}(A)$ ;
- (2)  $A\mathfrak{B}(S)$  is a closed  $W$ -symmetric manifold in  $\mathfrak{M}$  with  $W$ -deficiency index  $(m+p, m)$  or  $(m, m+p)$  according as  $S$  has the deficiency index  $(0, p)$  or  $(p, 0)$ ;
- (3) if  $\mathfrak{R}$  is an arbitrary  $W$ -symmetric manifold in  $\mathfrak{M}$  such that  $\mathfrak{R} \supset A\mathfrak{B}(S)$ , the boundary condition  $Af \in \mathfrak{R}$  is degenerate.

Again we introduce the transformation  $B$  used in the proof of Theorem 4.13. Then, in accordance with Lemma 4.4, there exists a maximal  $Q$ -symmetric manifold  $\mathfrak{U}$  in  $\mathfrak{D}(B) = \mathfrak{D}(A_1)$  such that  $B\mathfrak{U}$  is closed and has the  $Q$ -deficiency index  $(m+p, m)$  or  $(m, m+p)$  according as  $(0, p)$  or  $(p, 0)$  is the  $Q$ -deficiency index of  $\mathfrak{U}$ . Thus  $\mathfrak{B} + \mathfrak{U}$  is the graph of a maximal symmetric

extension  $S$  of  $H$  and the deficiency index of  $S$  is the  $Q$ -deficiency index of  $\mathfrak{U}$ , by Theorem 4.2; moreover,  $\mathfrak{B}(S) = \mathfrak{B} + \mathfrak{U} \subset \mathfrak{D}(A)$ .

Now let  $X$  be the isometric transformation from  $\mathfrak{B}^+ + \mathfrak{B}^-$  to  $\mathfrak{M}$  such that  $A \equiv XB$ . Then, as we know,  $X\mathfrak{B}^+ = \mathfrak{M}^-$ ,  $X\mathfrak{B}^- = \mathfrak{M}^+$ , and  $W \equiv XQX^{-1}$ . Therefore  $XB\mathfrak{U} = A\mathfrak{U} = A\mathfrak{B}(S)$  is closed and has the  $W$ -deficiency index  $(m, m+p)$  or  $(m+p, m)$  according as  $(m+p, m)$  or  $(m, m+p)$  is the  $Q$ -deficiency index of  $B\mathfrak{N}$ , and consequently according as  $(p, 0)$  or  $(0, p)$  is the deficiency index of  $S$ . Thus  $S$  has the properties (1) and (2) of the theorem.

Now let  $\mathfrak{N}$  be a  $W$ -symmetric manifold in  $\mathfrak{M}$  such that  $A\mathfrak{B}(S) \subset \mathfrak{N}$ . Then, clearly  $H(\mathfrak{N}) \supseteq S$  and, since  $S$  is maximal symmetric and  $H(\mathfrak{N})$  symmetric, this is possible if and only if  $H(\mathfrak{N}) \equiv S$ . Thus  $\mathfrak{N}(A) \cdot \mathfrak{N} = A\mathfrak{B}(S)$  while  $A\mathfrak{B}(S)$  is closed, and therefore the boundary condition  $Af \in \mathfrak{N}$  is obviously degenerate, so that  $S$  satisfies (3).

Finally, to conclude this section, we prove a theorem which establishes unequivocally that the dependence on the operator  $A$  of the conditions given in Theorems 4.8, 4.9, and 4.11 is not in any sense superficial.

**THEOREM 4.17.** *Let  $\mathfrak{M}$  be an arbitrary Hilbert space, and let  $\mathfrak{M}^+$  and  $\mathfrak{M}^-$  be Hilbert spaces in  $\mathfrak{M}$  such that  $\mathfrak{M} \ominus \mathfrak{M}^+ = \mathfrak{M}^-$ . Let  $W$  be the linear transformation in  $\mathfrak{M}$  such that  $W = iI$  on  $\mathfrak{M}^+$ ,  $W = -iI$  on  $\mathfrak{M}^-$ . Let  $H$  be a closed linear symmetric transformation in a Hilbert space  $\mathfrak{S}$ , with deficiency index  $(\aleph_0, \aleph_0)$ .*

*Then, if  $V$  is an arbitrary closed isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$ , there exists a reduction operator  $C$  for  $H^*$ , with range-space  $\mathfrak{M}$  and associated unitary transformation  $W$  in  $\mathfrak{M}$ , such that the boundary condition  $Cf \in \mathfrak{R}(I - V)$  is nondegenerate and defines a linear symmetric extension of  $H$  which is not maximal and does not have a maximal closure. The operator  $C$  can be constructed so as to be either of type I or of type II.*

*If  $\mathfrak{D}(V)$  has the dimension number  $\aleph_0$ , there exists a second reduction operator  $D$  for  $H^*$ , with range-space  $\mathfrak{M}$  and associated unitary transformation  $W$  in  $\mathfrak{M}$ , such that the boundary condition  $Df \in \mathfrak{R}(I - V)$  is nondegenerate and defines a maximal symmetric extension of  $H$ . Here  $D$  can be chosen either of type I or of type II.*

Our proof of this theorem requires that we first review the proof of Lemma 4.4, in order to take account of certain facts which are implicit there and which, for present purposes, must be formally stated. We observe first that the manifold  $\mathfrak{N}$  specified in the lemma is in the proof determined as an arbitrary maximal  $Q$ -symmetric extension of a  $Q$ -symmetric manifold  $\mathfrak{N}_1$  in  $\mathfrak{E} \ominus \mathfrak{P}$ , where  $\mathfrak{P}$  is the manifold in which  $Q = I$ ; and that  $\mathfrak{N}_1$  is hypermaximal  $Q_1$ -symmetric in  $\mathfrak{E} \ominus \mathfrak{P}$ , where  $Q_1$  is the contraction of  $Q$  with domain  $\mathfrak{E} \ominus \mathfrak{P}$ . Accordingly, if the transformation  $D$  of Lemma 4.4 is such that  $\mathfrak{P} \cdot \mathfrak{E}^+ = \mathfrak{D}$

while  $\mathfrak{B} \cdot \mathfrak{E}^-$  has dimension number  $p$ , then  $\mathfrak{N}$  has  $Q$ -deficiency index  $(0, p)$ . Similarly, if  $\mathfrak{B} \cdot \mathfrak{E}^- = \mathfrak{D}$ , while  $\mathfrak{B} \cdot \mathfrak{E}^+$  has the dimension number  $p$ , then  $\mathfrak{N}$  has the  $Q$ -deficiency index  $(p, 0)$ . Furthermore, if  $\mathfrak{B} \cdot \mathfrak{E}^+$  and  $\mathfrak{B} \cdot \mathfrak{E}^-$  both have the dimension number  $\aleph_0$ , then  $\mathfrak{N}$  can be chosen with arbitrary  $Q$ -deficiency index  $(0, p)$  or  $(p, 0)$ . These facts we hold in reserve for application later.

Now let  $V_1$  be an arbitrary maximal isometric extension from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  of the transformation  $V$  of the theorem to be proved, and let  $V_1$  have  $W$ -deficiency index  $(0, p)$   $((p, 0))$ ; further, let  $A_1$  be a nonnegative definite self-adjoint transformation in  $\mathfrak{B}^+ + \mathfrak{B}^-$ , such that  $A_1 \equiv QA_1^{-1}Q^{-1}$ , where  $\mathfrak{B}^+$ ,  $\mathfrak{B}^-$ , and  $Q$  have the usual meanings, with reference to the transformation  $H$  of Theorem 4.17. Then, by Theorem 3.1 the linear transformation  $A$  with domain  $\mathfrak{B}(H) + \mathfrak{D}(A_1)$ , which is equal to 0 on  $\mathfrak{B}(H)$  and to  $A_1$  on  $\mathfrak{D}(A_1)$  is a reduction operator for  $H^*$ . We consider two distinct cases: (1)  $A$  is unbounded of type I with characteristic index  $(p, 0)$   $((0, p))$ ; (2)  $A$  is unbounded of type II and the intersections of  $\mathfrak{B}^+$  and  $\mathfrak{B}^-$  with the manifold  $\mathfrak{P}$  on which  $A_1 = I$  are both Hilbert spaces. Since  $H$  has deficiency index  $(\aleph_0, \aleph_0)$ , the results of Chapter III, §2 assure the existence of reduction operators  $A$  satisfying either requirement.

We now apply Lemma 4.4 to establish the existence in  $\mathfrak{D}(A_1^{-1}) = \mathfrak{N}(A_1)$  of a maximal  $Q$ -symmetric manifold  $\mathfrak{N}$  such that  $A_1^{-1}\mathfrak{N}$  is closed linear  $Q$ -symmetric but not maximal; here we of course employ the relation  $A_1^{-1} \equiv QA_1Q^{-1}$ . Moreover, taking account of the paragraph immediately following Theorem 4.17, we see that under case (1) of the preceding paragraph,  $\mathfrak{N}$  has the  $Q$ -deficiency index  $(p, 0)$   $((0, p))$ , while under case (2),  $\mathfrak{N}$  can be chosen with the same  $Q$ -deficiency index. In the latter case, we assume that  $\mathfrak{N}$  is so chosen.

Applying Theorem 2.2, we next introduce the isometric transformation  $U$  from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that  $\mathfrak{N} = \mathfrak{N}(I - U)$ . Then  $U$  has, under either of the cases (1) and (2), the  $Q$ -deficiency index  $(p, 0)$   $((0, p))$ . Thus, by definition,  $\mathfrak{B}^- \ominus \mathfrak{N}(U)$  and  $\mathfrak{B}^+ \ominus \mathfrak{D}(U)$  have the same dimension numbers as  $\mathfrak{M}^+ \ominus \mathfrak{D}(V_1)$  and  $\mathfrak{M}^- \ominus \mathfrak{N}(V_1)$ , respectively. Consequently,  $\mathfrak{D}(U)$ ,  $\mathfrak{N}(U)$ ,  $\mathfrak{D}(V_1)$ , and  $\mathfrak{N}(V_1)$  being Hilbert spaces, we can define an isometric transformation  $Y$  with domain  $\mathfrak{B}^+ + \mathfrak{B}^-$  and range  $\mathfrak{M}$  such that  $Y\mathfrak{B}^- = \mathfrak{M}^+$ ,  $Y\mathfrak{B}^+ = \mathfrak{M}^-$ ,  $Y\mathfrak{N}(U) = \mathfrak{D}(V_1)$ , and  $Y\mathfrak{D}(U) = \mathfrak{N}(V_1)$ . Thus  $Y\mathfrak{N} = \mathfrak{N}(I - V_1)$ .

But, by Theorem 3.3,  $C \equiv YA$  is a reduction operator for  $H^*$ , with range-space  $\mathfrak{M}$ , and associated transformation  $W$  in  $\mathfrak{M}$ , since  $YQ^{-1}Y^{-1}$  is evidently identical with  $W$ . Furthermore, the condition  $Cf \in \mathfrak{N}(I - V_1)$  clearly defines the same symmetric extension of  $H$  as the condition  $Af \in \mathfrak{N}(I - U)$ , that is, the extension  $S$  whose graph is  $\mathfrak{B} + A_1^{-1}\mathfrak{N}$ ; and  $S$  is clearly closed linear but not maximal, since  $A_1^{-1}\mathfrak{N}$  is a closed linear manifold and not maximal  $Q$ -symmetric. The argument here is, of course, essentially that used to prove Theorem

4.15. Hence, the condition  $Cf \in \mathfrak{R}(I - V)$  defines an extension  $T$  of  $H$  which does not have a maximal symmetric closure, since we obviously have  $\tilde{T} \subseteq S$ . Moreover, since  $\mathfrak{R}(I - U) \subset \mathfrak{R}(A)$ , we have

$$\mathfrak{R}(I - V) \subseteq \mathfrak{R}(I - V_1) \subset \mathfrak{R}(C),$$

so that the condition  $Cf \in \mathfrak{R}(I - V)$  is nondegenerate. Finally, since we have shown that  $A$  can be either of type I or type II,  $C$  can be of either type also, by Theorem 3.16. Thus the first portion of the theorem is completely proved.

We turn now to the second, and denote by  $(n, q)$  the  $W$ -deficiency index of  $V$ . For purposes of simplification, we assume  $n \geq q$ ; the alternative possibility can be discussed along lines entirely similar to those which we pursue in this case.

Setting  $p = n - q$ , we apply Theorems 3.5 and 3.6 to construct a reduction operator  $A$  for  $H^*$  such that  $A_1$  is a nonnegative definite self-adjoint transformation in  $\mathfrak{B}^+ + \mathfrak{B}^-$  and such that the manifold on which  $A_1 = I$  is a linear manifold in  $\mathfrak{B}^-$  with dimension number  $p$ . We then apply Lemma 4.4 and the facts stated in the paragraph immediately following Theorem 4.17 to determine a maximal  $Q$ -symmetric manifold  $\mathfrak{R}$  in  $\mathfrak{B}^+ + \mathfrak{B}^-$  such that  $A_1\mathfrak{R}$  is closed linear  $Q$ -symmetric with the  $Q$ -deficiency index  $(q, p + q)$ .

Next, we introduce the isometric transformation  $U$  from  $\mathfrak{B}^+$  to  $\mathfrak{B}^-$  such that  $A_1\mathfrak{R} = \mathfrak{R}(I - U)$ . Then, by definition,  $\mathfrak{B}^+ \ominus \mathfrak{D}(U)$  and  $\mathfrak{B}^- \ominus \mathfrak{R}(U)$  have the dimension numbers  $q$  and  $p + q = n$ , respectively. Consequently, there exists an isometric transformation  $Y$  from  $\mathfrak{B}^+ + \mathfrak{B}^-$  to  $\mathfrak{M}$  such that

$$Y\mathfrak{B}^+ = \mathfrak{M}^-, \quad Y\mathfrak{B}^- = \mathfrak{M}^+, \quad Y\mathfrak{D}(U) = \mathfrak{R}(V), \quad Y\mathfrak{R}(U) = \mathfrak{D}(V).$$

Thus  $Y\mathfrak{R}(I - U) = \mathfrak{R}(I - V)$ .

But, again by Theorem 3.3,  $D \equiv YA$  is a reduction operator for  $H^*$ , with range-space  $\mathfrak{M}$  and associated transformation  $W$  in  $\mathfrak{M}$ ; furthermore,  $\mathfrak{R}(I - V)$  clearly belongs to  $\mathfrak{R}(D)$ , so that the boundary condition  $Df \in \mathfrak{R}(I - V)$  is nondegenerate. Therefore, since this boundary condition obviously defines a symmetric extension  $S$  of  $H$  with graph  $\mathfrak{B} + \mathfrak{R}$  and  $\mathfrak{R}$  is maximal  $Q$ -symmetric, the first assertion of the second paragraph of the theorem now follows at once from Theorem 4.2. Finally, since it is evident from Theorem 3.13 that  $A$  can be chosen either of type I or of type II, subject to our previous conditions, we conclude from Theorem 3.16 that  $D$  can be of either type.

**6. A special class of boundary conditions.** From Theorem 4.17 we of course conclude that the general result of Theorem 4.8 admits of no effective simplification. Consequently, in dealing with any particular concrete unbounded reduction operator  $A$ , one would naturally look for special properties of  $A$  which might lead to less general but more readily applicable results. In

this connection, both for its illustrative value and its usefulness in connection with certain applications of the theory here developed, we now consider a specialization of the general situation which we have heretofore been studying and subject it to study by methods of a quite different sort from those employed in §4.

We require a preliminary lemma.

LEMMA 4.5. *Let  $R$  be a transformation in a Hilbert space  $\mathfrak{H}$ , with at least one point  $\mu$  in its resolvent set, and let  $(R - \mu I)^{-1}$  be totally continuous. Then, if  $T$  is an arbitrary bounded transformation in  $\mathfrak{H}$  and  $\lambda$  belongs to the resolvent set of  $R + T$ ,  $(R + T - \lambda I)^{-1}$  is totally continuous. In particular,  $(R - \lambda I)^{-1}$  is totally continuous for every point  $\lambda$  in the resolvent set of  $R$ .*

We consider an arbitrary set  $\mathfrak{U}$  in  $\mathfrak{R}(R + T - \lambda I) = \mathfrak{H}$ ,  $\lambda$  being in the resolvent set of  $R + T$ . Then, if  $\mu$  is in the resolvent set of  $R$ ,

$$\mathfrak{D}(R - \mu I) = \mathfrak{D}(R + T - \lambda I) = \mathfrak{R}((R + T - \lambda I)^{-1})$$

and

$$\begin{aligned} (R + T - \lambda I)^{-1}\mathfrak{U} &= (R - \mu I)^{-1}(R - \mu I)(R + T - \lambda I)^{-1}\mathfrak{U} \\ &= (R - \mu I)^{-1}[(-T + (\lambda - \mu)I)(R + T - \lambda I)^{-1} + I]\mathfrak{U}. \end{aligned}$$

Now suppose  $\mathfrak{U}$  is a bounded set. Then, since

$$[(-T + (\lambda - \mu)I)(R + T - \lambda I)^{-1}]$$

is a bounded transformation,

$$[(-T + (\lambda - \mu)I)(R + T - \lambda I)^{-1} + I]\mathfrak{U}$$

is a bounded set. Thus, if  $(R - \mu I)^{-1}$  is totally continuous, it follows from the above equation that  $(R + T - \lambda I)^{-1}\mathfrak{U}$  is compact. Thus  $(R + T - \lambda I)^{-1}$  is totally continuous, as we wished to show. Since, in particular, we can take  $T = 0$ , the last statement of the lemma follows immediately, and the proof is complete.

THEOREM 4.18. *Let  $H$  have the deficiency index  $(n, n)$ , and let  $A$  be a reduction operator of the sort described in Theorem 3.17, with range-space  $\mathfrak{H} \oplus \mathfrak{H}$ . Let  $S$  be the self-adjoint transformation in  $\mathfrak{S} \oplus \mathfrak{H}$  which corresponds to  $A$  in accordance with Theorem 3.17. Let  $\mathfrak{D}_i^*$  be the set of elements  $f$  of  $\mathfrak{D}^*$  such that  $\{f, H^*f\}$  is in  $\mathfrak{D}(A)$ , and let  $N$  and  $M$  be the operators, each with domain  $\mathfrak{D}_i^*$  and range in  $\mathfrak{H}$ , such that  $Nf = h$  and  $Mf = k$  if and only if  $A\{f, H^*f\} = \{h, k\}$ . Then  $N$  and  $M$  are linear transformations.*

*If  $S$  has a totally continuous resolvent and  $L$  is an arbitrary bounded self-adjoint transformation in  $\mathfrak{H}$ , then the boundary condition  $Af \in \mathfrak{B}(L)$ , that is, the requirement  $Mf = LNf$ , defines a self-adjoint extension of  $H$ .*

The assertion concerning  $M$  and  $N$  is evident.

To prove the remainder of the theorem, we denote by  $L_1$  the transformation with domain  $\mathfrak{S} \oplus \mathfrak{L}$  which is equal to  $O$  in  $\mathfrak{S}$  and to  $L$  in  $\mathfrak{L}$ , and assume that  $S$  has a totally continuous resolvent. Then  $L_1$  is clearly a bounded self-adjoint transformation in  $\mathfrak{S} \oplus \mathfrak{L}$  and it is readily shown that  $S - L_1$  is self-adjoint. Moreover, by Lemma 4.5,  $S - L_1$  has a totally continuous resolvent; and it can therefore be shown that the range of  $S - L_1$  is the orthogonal complement of its manifold of zeros.

Therefore, in particular, the equations  $H^*f = f^*$ ,  $Mf - LNf = 0$  have a solution  $f$  in  $\mathfrak{D}_1^*$  for every  $f^*$  in  $\mathfrak{S} \ominus \mathfrak{U}$ , where  $\mathfrak{U}$  is the manifold of elements  $g$  of  $\mathfrak{D}_1^*$  such that  $H^*g = 0$ ,  $Mf - LNf = 0$ . But, by definition, every element  $f$  of  $\mathfrak{D}_1^*$  such that  $Mf - LNf = 0$  belongs to the domain of  $H(\mathfrak{B}(L))$ , where we have reverted to the notation of Definition 1.2; and  $H(\mathfrak{B}(L))$  is symmetric, by Theorems 1.5 and 2.7. Moreover, from the result of the preceding paragraph, the range of  $H(\mathfrak{B}(L))$  is the orthogonal complement of its manifold of zeros, which we have denoted by  $\mathfrak{U}$ . Thus  $\mathfrak{U}$  and  $\mathfrak{S} \ominus \mathfrak{U}$  reduce  $H(\mathfrak{B}(L))$ ; in  $\mathfrak{U}$ ,  $H(\mathfrak{B}(L))$  induces the self-adjoint transformation  $O$ ; and, in  $\mathfrak{S} \ominus \mathfrak{U}$ ,  $H(\mathfrak{B}(L))$  induces a transformation whose range is  $\mathfrak{S} \ominus \mathfrak{U}$  and which is therefore self-adjoint. Thus  $H(\mathfrak{B}(L))$  itself is self-adjoint in  $\mathfrak{S}$  and the proof is complete.

The reader may observe that  $H(\mathfrak{B}(L))$ , being related very simply to a contraction of  $S - L_1$ , has itself a totally continuous resolvent and thus a pure point spectrum; we refrain, however, from discussing questions of this sort here, but rather reserve them for separate consideration elsewhere.

The hypothesis that  $S$  has a totally continuous resolvent which appears in Theorem 4.18 is more restrictive than is, in fact, necessary. For it imposes a restriction on the behavior of the transformation  $H$  itself, and this can evidently have no effect on the behavior of the transformation  $A$ . However, both the statement and the proof of the more general theorem which is possible are considerably more involved than those of the one which we have given; and, in all of the realizations which we have investigated, Theorem 4.18 applies whenever the more general result does. (In particular, this remark applies to Example 4 of the first chapter; the reduction operator  $A$  described there satisfies the hypothesis of Theorem 4.18.)

We have therefore refrained from giving the more general result here and also from investigating further along lines suggested by Theorem 4.18 and the fact already noted that every equivalence class of reduction operators contains, when  $H$  has deficiency index  $(n, n)$ , operators of the sort described in Theorem 3.17: investigations in this direction and other similar ones must be guided to some extent by the nature of possible applications, of which we have at present studied only a few.

**7. Real boundary conditions.** If the transformation  $H$  is real with respect to a conjugation  $J_0$  in  $\mathfrak{S}$ , then  $H^*$  is real with respect to  $J_0$  also and  $H$  has deficiency index  $(n, n)$ .<sup>†</sup> In such a situation it is frequently important to determine those maximal symmetric extensions of  $H$  which are real with respect to  $J_0$  and which are consequently self-adjoint. For the special case that  $A$  is the reduction operator of Theorem 2.9, this problem has a simple solution which is known;<sup>‡</sup> we consider here a more general case.

**THEOREM 4.19.** *Let  $H$  be real with respect to a conjugation  $J_0$  in  $\mathfrak{S}$ , and let  $J$  be the transformation in  $\mathfrak{S} \oplus \mathfrak{S}$  which takes  $\{f, g\}$  into  $\{J_0 f, J_0 g\}$ . Then, in accordance with Theorem 2.14,  $J$  permutes with  $Q$  in  $\mathfrak{B}^+ + \mathfrak{B}^-$  and  $\mathfrak{B}^*$  is real with respect to  $J$ .*

*Let  $A$  be a reduction operator for  $H^*$ , let  $\mathfrak{D}(A)$  be real with respect to  $J$ , and let  $J_1$  be a conjugation in the range-space  $\mathfrak{M}$  of  $A$  such that  $AJ\{f, H^*f\} = J_1 A\{f, H^*f\}$  for every  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$ . Then  $J_1$  permutes with  $W$ .*

*Let  $V$  be an arbitrary closed isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that the boundary condition  $Af \in \mathfrak{R}(I - V)$  is nondegenerate. Then  $H(V)$  is real with respect to  $J_0$  if and only if  $\mathfrak{R}(I - V)$  is real with respect to  $J_1$  or, equivalently, if and only if  $V \equiv J_1 V^{-1} J_1$ . Thus, if  $V \equiv J_1 V^{-1} J_1$ , then  $\tilde{H}(V)$  has deficiency index  $(n, n)$ ; consequently  $H(V)$  is self-adjoint if it is maximal symmetric, essentially self-adjoint if  $\tilde{H}(V)$  is maximal symmetric.*

As we have noted in the statement of the theorem, the assertions of the first paragraph are consequences of Theorem 2.14.

We now prove that  $J_1$  permutes with  $W$ . Let  $\{f, H^*f\}, \{g, H^*g\}$  be arbitrary elements of  $\mathfrak{D}(A_1)$ . Then

$$\begin{aligned} (J_1 A_1 \{f, H^*f\}, J_1 W A_1 \{g, H^*g\}) &= \overline{(A_1 \{f, H^*f\}, W A_1 \{g, H^*g\})} \\ &= -(\{f, H^*f\}, Q\{g, H^*g\}), \end{aligned}$$

where the last inner product is formed in  $\mathfrak{S} \oplus \mathfrak{S}$ . But

$$\overline{(\{f, H^*f\}, Q\{g, H^*g\})} = (J\{f, H^*f\}, QJ\{g, H^*g\}),$$

since  $J$  permutes with  $Q$ ; and

$$\begin{aligned} (J\{f, H^*f\}, QJ\{g, H^*g\}) &= -(A_1 J\{f, H^*f\}, W A_1 J\{g, H^*g\}) \\ &= -(J_1 A_1 \{f, H^*f\}, W J_1 A_1 \{g, H^*g\}). \end{aligned}$$

Thus we have

$$(J_1 A_1 \{f, H^*f\}, J_1 W A_1 \{g, H^*g\}) = (J_1 A_1 \{f, H^*f\}, W J_1 A_1 \{g, H^*g\})$$

<sup>†</sup> Stone, Theorems 9.13 and 9.14.

<sup>‡</sup> Stone, pp. 362-364.

for all  $\{f, H^*f\}, \{g, H^*g\}$  in  $\mathfrak{D}(A_1)$ . But  $J_1A_1\mathfrak{D}(A_1)$  is dense in  $\mathfrak{M}$  and hence

$$J_1WA_1\{g, H^*g\} = WJ_1A_1\{g, H^*g\}$$

for all  $\{g, H^*g\}$  in  $\mathfrak{D}(A)$ . Therefore  $J_1W = WJ_1$  on  $\mathfrak{R}(A)$ . But  $\overline{\mathfrak{R}(A)} = \mathfrak{M}$ , while  $J_1W$  and  $WJ_1$  are continuous. Consequently  $J_1W \equiv WJ_1$ , as we wished to prove.

Now let  $V$  be the transformation described in the last paragraph of the theorem. We note first that  $H(V)$  is real with respect to  $J_0$  if and only if its graph is real with respect to  $J$ ; furthermore,  $\mathfrak{B}$  is clearly real with respect to  $J$ , since  $H$  is real with respect to  $J_0$ . Hence, setting

$$\mathfrak{B}_1(H(V)) = \mathfrak{B}(H(V)) \cdot (\mathfrak{B}^+ + \mathfrak{B}^-),$$

we conclude that  $H(V)$  is real with respect to  $J_0$  if and only if  $\mathfrak{B}_1(H(V))$  is real with respect to  $J$ ; and thus, by Theorem 2.12, if and only if  $J\mathfrak{B}_1(H(V)) = \mathfrak{B}_1(H(V))$ .

We shall now show that the latter equation holds if and only if  $\mathfrak{R}(I-V)$  is real with respect to  $J_1$ , beginning with the observation that  $J_1A_1\mathfrak{B}_1(H(V)) = A_1J\mathfrak{B}_1(H(V))$ . Hence, if  $J\mathfrak{B}_1(H(V)) = \mathfrak{B}_1(H(V))$ , we must have  $J_1A_1\mathfrak{B}_1(H(V)) = A_1\mathfrak{B}_1(H(V))$ , and the latter implies that  $A_1\mathfrak{B}_1(H(V))$  is real with respect to  $J_1$ , by Theorem 2.12. Hence, since  $\mathfrak{R} = \mathfrak{R}(I-V)$  is the closure of  $A_1\mathfrak{B}_1(H(V))$ ,  $\mathfrak{R}$  is real with respect to  $J_1$ , again by Theorem 2.12. On the other hand, suppose  $\mathfrak{R} = \mathfrak{R}(I-V)$  is real with respect to  $J_1$ . Then, since  $J_1A\{f, H^*f\}$  is in  $\mathfrak{R}(A)$  for every  $\{f, H^*f\}$  in  $\mathfrak{D}(A)$ ,  $\mathfrak{R}(A)$  is real with respect to  $J_1$  and therefore  $\mathfrak{R} \cdot \mathfrak{R}(A)$  is also. But

$$A_1^{-1}[\mathfrak{R} \cdot \mathfrak{R}(A)] = \mathfrak{B}_1(H(V))$$

and

$$A_1^{-1}[\mathfrak{R} \cdot \mathfrak{R}(A)] = A_1^{-1}J_1[\mathfrak{R} \cdot \mathfrak{R}(A)] = JA_1^{-1}[\mathfrak{R} \cdot \mathfrak{R}(A)].$$

Consequently  $J\mathfrak{B}_1(H(V)) = \mathfrak{B}_1(H(V))$ .

Thus, combining the results of the two preceding paragraphs, we conclude that  $H(V)$  is real with respect to  $J_0$  if and only if  $\mathfrak{R} = \mathfrak{R}(I-V)$  is real with respect to  $J_1$ ; and, as we have already shown that  $W$  permutes with  $J_1$ , the equivalence of the condition that  $\mathfrak{R}$  be real with respect to  $J$  and the condition  $V \equiv JV^{-1}J$  follows from Theorem 2.13.

We are thus left to prove only the assertions of the final sentence in the theorem. This, however, requires only the observation that  $\tilde{H}(V)$  is evidently real with respect to  $J_0$  when  $H(V)$  is; the statements in question then follow from the results already established and Theorem 9.14 of the book of Stone previously cited.

We point out that the hypothesis of the existence of the conjugation  $J_1$  is necessary and does not follow from the reality of  $H$  and  $H^*$ ; the hypothesis is, however, in accord with the situation which exists in the theory of differential operators. If such an assumption is not introduced, apparently little of interest can be said concerning those boundary conditions which define real extensions of  $H$ .

Theorem 4.19 has, of course, a special interpretation for the boundary conditions considered in Theorem 4.18; we leave the formulation of this to the reader.

**8. Formulation of boundary conditions.** In developing our theory we have found it necessary only to use the representations  $Af \in \mathfrak{N}$ , or  $Af \in \mathfrak{N}(I - V)$  for the boundary conditions under consideration. In the applications of the theory, however, different statements of the conditions may be found convenient. We now discuss briefly some of these. We consider an arbitrary nondegenerate boundary condition  $Af \in \mathfrak{N}$ ,  $\mathfrak{N}$  being closed linear  $W$ -symmetric.

If  $\{\phi_n\}$ ,  $n=1, 2, \dots$ , is an arbitrary sequence which determines the closed linear manifold  $\mathfrak{M} \ominus \mathfrak{N}$ , the condition  $Af \in \mathfrak{N}$  is evidently equivalent to the conditions  $(Af, \phi_n) = 0$ ,  $n=1, 2, \dots$ . Now let  $\{\psi_m\}$  be a complete orthonormal set in  $\mathfrak{M}$ , and let  $\phi_n = \sum_m \bar{a}_{mn} \psi_m$ ,  $n=1, 2, \dots$ ;  $Af = \sum_m b_m(f) \psi_m$ , for each  $Af$  in  $\mathfrak{N}$ . Then  $(Af, \phi_n) = \sum_m a_{mn} b_m(f)$ . Accordingly, since the sequences  $\{b_m(f)\}$  are evidently in one-to-one correspondence with the elements  $Af$  in  $\mathfrak{N}$ , the condition  $Af \in \mathfrak{N}$  is equivalent to the conditions  $\sum_m a_{mn} b_m(f) = 0$ ,  $n=1, 2, \dots$ . If  $\mathfrak{M}$  is a Hilbert space, there is, in general, an infinite number of these equations; if  $\mathfrak{M}$  is unitary, there is only a finite number.

We may point out that the form of representation of the boundary conditions generally used in the theory of ordinary differential equations is precisely of the sort just described.

Another simple representation of the boundary conditions is a parametric one. If  $\mathfrak{N}$  is  $W$ -symmetric in  $\mathfrak{M}$  and  $V$  is the isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $\mathfrak{N} = \mathfrak{N}(I - V)$ , we may regard the manifold  $\mathfrak{D}(V)$  as a space of parameters and define the domain of  $H(V)$  as that set of elements  $f$  of  $\mathfrak{D}^*$  such that  $Af$  is defined and satisfies the equation  $Af = h - Vh$  for some  $h$  in  $\mathfrak{D}(V)$ . In this connection, we observe that we are at liberty to choose coordinates in  $\mathfrak{M}^+$  and in  $\mathfrak{M}$  as we please.

Finally, we consider the special case where  $A$  is a reduction operator of the kind described in Theorem 3.17. Here, the range of  $A$  is in a space  $\mathfrak{E} \oplus \mathfrak{E}$ , and we can set  $Af = \{Nf, Mf\}$ , where  $Nf$  and  $Mf$  have the same meanings as in Theorem 4.18. In this case, if  $\mathfrak{N}$  is a closed  $W$ -symmetric manifold in  $\mathfrak{E} \oplus \mathfrak{E}$ , then by Theorem 2.10, there exists a uniquely determined closed linear

manifold  $\mathfrak{U}$  in  $\mathfrak{E}$  and a uniquely determined Hermitian transformation  $L$  in  $\mathfrak{E} \ominus \mathfrak{U}$ , such that the condition  $Af \in \mathfrak{N}$  is equivalent to the condition  $Mf = LNf + h$ , where  $h$  is in  $\mathfrak{U}$ . In particular, the boundary conditions of potential theory are of this form. Moreover, M. Morse has found essentially this form of representation useful in connection with comparison theorems for systems of ordinary differential equations of the second order.† In the terminology of Morse, the space  $\mathfrak{E} \ominus \mathfrak{U}$  is the "accessory end-plane," and the form  $(k, Lk)$  the "accessory end-form."

**9. Homogeneous boundary value problems.** The theory which we have developed (particularly the results of Chapter IV, §4) provides information concerning the solvability of the equation  $H^*f - \lambda f = f^*$  under a restriction  $Af \in \mathfrak{N}$ , where  $\mathfrak{N}$  is a  $W$ -symmetric manifold in  $\mathfrak{M}$ . For example, if the boundary condition  $Af \in \mathfrak{N}$  defines a self-adjoint extension of  $H$ , and if  $f^*$  is an arbitrary element of  $\mathfrak{S}$ ,  $\lambda$  an arbitrary complex such that  $\Im(\lambda) \neq 0$ , then there exists one and only one element  $f$  of  $\mathfrak{D}^*$  such that  $H^*f - \lambda f = f^*$ ,  $Af \in \mathfrak{N}$ .

The theory of differential equations suggests quite naturally that we should consider, as well as questions of the above sort, problems of the following kind: given a linear manifold  $\mathfrak{N}$  in  $\mathfrak{M}$ , to determine an element  $f$  of  $\mathfrak{D}^*$  such that  $H^*f - \lambda f = 0$ ,  $E_{\mathfrak{N}}Af = h$ , where  $\lambda$  is a complex number and  $h$  a pre-assigned element of  $\mathfrak{N}$ . We bring this paper to its end with a few observations concerning such problems.

**THEOREM 4.20.** *Let  $V$  be an arbitrary closed isometric transformation from  $\mathfrak{M}^+$  to  $\mathfrak{M}^-$  such that  $H(V)$  is a maximal symmetric extension of  $H$ . Let  $\mathfrak{N} = \mathfrak{N}(I - V)$ , and let  $\lambda$  belong to the resolvent set of  $H(V)$ . Then, for every element  $h$  of an everywhere dense linear manifold in  $\mathfrak{P} = \mathfrak{M} \ominus \mathfrak{N}$ , there exists a solution  $f$  of the equation  $H^*f - \lambda f = 0$  such that  $Af$  is defined and  $E_{\mathfrak{P}}Af = h$ ; if  $\mathfrak{P}$  belongs to the range of  $A$  in particular, if  $A$  is bounded, then a solution  $f$  exists for every  $h$  in  $\mathfrak{P}$ . Moreover,  $f$  is uniquely determined by  $h$ .*

To prove this theorem, we note first that if  $\lambda$  is in the resolvent set of  $H(V)$ , every element  $g$  of  $\mathfrak{D}^*$  has a unique resolution of the form  $g = f_1 + f$ , where  $f_1$  belongs to  $\mathfrak{D}(H(V))$  and  $f$  is a solution of the equation  $H^*f - \lambda f = 0$ ; to demonstrate this, we have only to set

$$f_1 = (H(V) - \lambda I)^{-1}(H^* - \lambda I)g, \quad f = g - f_1.$$

In particular, every element  $g$  such that  $Ag$  is defined can be written in this form and, for such an element, since  $Af_1$  is defined by hypothesis,  $Af$  is also. Furthermore, for every such  $g$ ,  $E_{\mathfrak{P}}Ag = E_{\mathfrak{P}}Af$ , since  $Af_1$  is in  $\mathfrak{N} = \mathfrak{M} \ominus \mathfrak{P}$ .

† M. Morse, *The Calculus of Variations in the Large*, American Mathematical Society Colloquium Publications, vol. 18, New York, 1934, chap. 4, especially pp. 83-89.

Therefore, the system  $H^*f - \lambda f = 0$ ,  $E_{\mathfrak{P}}Af = h$  has a solution  $f$  for every element  $h$  in  $E_{\mathfrak{P}}\mathfrak{R}(A)$ . But  $\mathfrak{R}(A)$  is dense in  $\mathfrak{M}$  and  $E_{\mathfrak{P}}\mathfrak{R}(A)$  is therefore a dense linear manifold in  $\mathfrak{P}$ . Moreover, if  $\mathfrak{P}$  belongs to  $\mathfrak{R}(A)$  (and this is necessarily true when  $A$  is bounded) we have  $E_{\mathfrak{P}}\mathfrak{R}(A) = \mathfrak{P}$ .

Thus it remains only to prove that the solution  $f$  is unique. Let us suppose that  $f_1$  and  $f_2$  are two solutions for the same element  $h$  of  $\mathfrak{P}$ . Then  $A(f_1 - f_2)$  belongs to  $\mathfrak{M} \ominus \mathfrak{P} = \mathfrak{N}$ . Thus  $f_1 - f_2$  is in the domain of  $H(V)$ . But

$$H(V)(f_1 - f_2) - \lambda(f_1 - f_2) = 0$$

and  $\lambda$  is in the resolvent set of  $H(V)$ . Hence  $f_1 = f_2$ , and  $f$  is uniquely determined by  $h$ , as we wished to show.

It is interesting to observe that suitably modified formulations of the Dirichlet and Neumann problems of potential theory can be described in terms of the abstract problem considered in Theorem 4.20. Furthermore, the so-called boundary value problem of the third kind is closely connected with a problem of this sort, as the following theorem suggests:

**THEOREM 4.21.** *Let  $A$  be a reduction operator of the sort described in Theorem 3.17 with range-space  $\mathfrak{Q} \oplus \mathfrak{Q}$ , and let  $M$  and  $N$  have the same meanings as in Theorem 4.18. Let  $L$  be a bounded self-adjoint transformation in  $\mathfrak{Q}$ , and let the boundary condition  $Af \in \mathfrak{B}(L)$  be nondegenerate and define a maximal symmetric extension of  $H$ . Then if  $\lambda$  is in the resolvent set of the extension so defined, the system*

$$H^*f - \lambda f = 0, \quad LNf - Mf = k$$

*has a unique solution  $f$  for every element  $k$  of an everywhere dense linear manifold in  $\mathfrak{Q}$ .*

According to Theorem 4.20, the system

$$H^*f - \lambda f = 0, \quad E_{\mathfrak{P}}\{Nf, Mf\} = \{r, s\}$$

has a unique solution  $f$  for every element  $\{r, s\}$  in a dense linear subset of  $\mathfrak{P}$ , where  $\mathfrak{P} = (\mathfrak{Q} \oplus \mathfrak{Q}) \ominus \mathfrak{B}(L)$ . Since  $L$  is self-adjoint, in the second equation above, we can set  $r = Lh$ ,  $s = -h$ . Then  $\{Nf, Mf\} = \{Lh, -h\} + \{t, Lt\}$ , where  $t$  is some element of  $\mathfrak{Q}$ . Thus

$$LNf = L^2h + Lt, \quad Mf = -h + Lt.$$

Therefore  $LNf - Mf = (L^2 + I)h$ . Thus the solution  $f$  of the system

$$H^*f - \lambda f = 0, \quad E_{\mathfrak{P}}\{Nf, Mf\} = \{Lh, -h\}$$

is also a solution of the system

$$H^*f - \lambda f = 0, \quad LNf - Mf = k,$$

with  $k = (L^2 + I)h$ . Conversely, it is easily shown that a solution  $f$  of the second system determines a solution of the first, again with  $k = (L^2 + I)h$ . Therefore, since the equation  $k = (L^2 + I)h$  is readily shown to determine a one-to-one linear bicontinuous correspondence between the elements  $k$  of  $\mathfrak{K}$  and the elements  $\{Lh, -h\}$  of  $\mathfrak{P}$ , the theorem follows.

If, in particular, the transformation  $S$  in  $\mathfrak{S} \oplus \mathfrak{K}$ , associated with  $A$  by Theorem 3.17, satisfies the hypothesis of the second paragraph of Theorem 4.18, we are able to attack questions of the sort considered in Theorem 4.21 by methods similar to those used in proving Theorem 4.18.

**THEOREM 4.22.** *Let  $A$ ,  $\mathfrak{K}$ ,  $M$ ,  $N$ , and  $S$  have the same meanings as in Theorem 4.18, and let  $S$  have a totally continuous resolvent. Let  $L$  be an arbitrary bounded self-adjoint transformation in  $\mathfrak{K}$ , and let  $T$  be the extension of  $H$  determined by the boundary condition  $Af \in \mathfrak{B}(L)$ ; that is to say, by the condition  $LNf = Mf$ . Let  $\lambda$  be a real number, and let  $(T - \lambda I)^{-1}$  exist. Then the system*

$$H^*f - \lambda f = 0, \quad Mf - LNf = h$$

*has one and only one solution  $f$  for every  $h$  in  $\mathfrak{K}$ .*

In view of Theorem 1.3, it is sufficient to prove the theorem for the case  $\lambda = 0$ .

Again we introduce the transformation  $L_1$  used in the proof of Theorem 4.18. Then, since  $S - L_1$  has a totally continuous resolvent, it is a simple task to show that the origin is either in its point spectrum or its resolvent set; we omit the details. Since  $T^{-1}$  exists, it is clear that the origin cannot belong to the point spectrum of  $S - L_1$ ; therefore, it belongs to the resolvent set. Accordingly, the equation  $(S - L_1)\{f, k\} = \{0, h\}$  has a unique solution  $\{f, k\}$  in  $\mathfrak{S} \oplus \mathfrak{K}$  for every  $h$  in  $\mathfrak{K}$ . Consequently, since  $f$  is then a solution of the system  $H^*f = 0$ ,  $Mf - LNf = h$ , the proof is complete.

To conclude, we suggest two simple generalizations, in different directions, of Theorem 4.22. First, if  $\lambda$  is a real number in the point spectrum of the transformation  $T$ , and  $\mathfrak{P} = N\mathfrak{U}$ , where  $\mathfrak{U}$  is the manifold of zeros of  $T$ , it can be shown that  $\mathfrak{P}$  has a finite dimension number and that the system

$$H^*f - \lambda f = 0, \quad Mf - LNf = h$$

has a solution  $f$  for every  $h$  in  $\mathfrak{K} \ominus \mathfrak{P}$ , but the solution is not unique. Second, Theorem 4.22 can be extended to cover the case that  $\lambda$  is not real, by taking account of the fact that the transformation in  $\mathfrak{S} \oplus \mathfrak{K}$  which takes  $\{f, Nf\}$  into  $\{H^*f - \lambda f, Mf\}$  is normal. We leave to the reader the proof of these assertions.

THE INSTITUTE FOR ADVANCED STUDY,  
PRINCETON, N. J.

# CONFORMAL GEODESICS\*

BY

AARON FIALKOW

1. **Introduction.** The totality of extremals in a Riemann space  $V_n$ † connected with a calculus of variations problem of the form

$$(1.1) \quad \delta \int F ds = 0,$$

where  $F$  is a point function and  $ds$  is the element of length of  $V_n$ , constitutes an important family of  $\infty^{2(n-1)}$  curves. Consider, for example, a conservative dynamical system for which neither the constraints nor the work function  $W$  involve the time. By the principle of least action the dynamical trajectories of a particle are the extremals of (1.1) with  $F = [2m(c+W)]^{1/2}$  where  $m$  and  $c$  are the mass of the particle and the energy constant, respectively. Again, if  $\nu$  is the index of refraction of an isotropic nonhomogeneous medium, the paths of light through this medium are the solutions of (1.1) with  $F = \nu$  in accordance with Fermat's principle. Finally, let  $V_n$  and  $\bar{V}_n$  be two conformal Riemann spaces so that  $d\bar{s} = e^{\sigma} ds$ . Then the images of the geodesics of  $\bar{V}_n$  in  $V_n$  are the extremals of (1.1) where  $F = e^{\sigma}$ .‡

As we are interested primarily in the last interpretation, following Schouten,§ we call any family of  $\infty^{2(n-1)}$  curves which is a solution of (1.1) a family of *conformal geodesics*. Of course, by a change of language, the theorems obtained have equal validity for the dynamical, optical, and other interpretations. The following topics are discussed and the corresponding questions answered in this paper:

I. A complete geometric characterization of the conformal geodesics of any Riemann space.

II. Additional special properties characteristic of conformal geodesics which are the images of the geodesics of a particular Riemann space (flat space, space of constant curvature, Einstein space).

\* Presented to the Society, February 26, 1938; received by the editors June 15, 1938.

† We denote an  $n$ -dimensional Riemann space, Einstein space, and space of constant curvature by  $V_n$ ,  $E_n$ , and  $S_n$ , respectively.

‡ These and other interpretations are discussed by E. Kasner, *Natural families of trajectories: conservative fields of force*, these Transactions, vol. 10 (1909), pp. 201-203. Also cf. L. P. Eisenhart, *Continuous Groups of Transformations*, 1933, pp. 277-280.

§ J. A. Schouten, *Über die Umkehrung eines Satzes von Lipschitz*, Nieuw Archief voor Wiskunde, vol. 15 (1928), pp. 97-102.

III. Some relations between the conformal geodesics of a Riemann space and of its subspaces.

IV. Some special geometric problems.

Other questions concerning conformal geodesics and other extremals (the theorems of Lipschitz, Thomson and Tait, Kneser, and their converses) have been investigated by Kasner, Lipke, Schouten, Blaschke, Douglas, LaPaz, and Radó.\* The first of the above topics was previously considered from the standpoint of dynamics and a solution obtained by Kasner† for the case of a euclidean space and by Lipke‡ for a Riemann space whose first fundamental form is positive definite. The characterization which they obtained is stated under more general conditions in Theorem 2.1 (or 2.2) and Theorem 3.2. The method which is used in the present paper differs from that hitherto employed.

The geometry of conformal geodesics is closely related to the more general investigation of the geometric properties of any curves or subspaces of  $V_n$  and  $\bar{V}_n$ , respectively, which correspond under the given conformal transformation. Somewhat similar studies of some phases of this problem have recently been made by a number of writers.§

#### I. GEOMETRIC CHARACTERIZATION

2. **Property one: the principal normal.** Let  $V_n$  and  $\bar{V}_n$  be two conformal  $n$ -dimensional Riemann spaces whose first fundamental forms are||

$$(2.1) \quad ds^2 = g_{ij}dx^i dx^j,$$

$$(2.2) \quad d\bar{s}^2 = \bar{g}_{ij}d\bar{x}^i d\bar{x}^j,$$

respectively, so that

$$(2.3) \quad d\bar{s} = e^\sigma ds.$$

\* For references to all of these writers cf. L. LaPaz and T. Radó, *On a converse of Kneser's transversality theorem*, *Annals of Mathematics*, (2), vol. 36 (1935), pp. 749-769.

† E. Kasner, loc. cit., pp. 201-219.

‡ J. Lipke, *Natural families of curves in a general curved space of  $n$  dimensions*, these *Transactions*, vol. 13 (1912), pp. 77-95.

§ S. Sasaki, *Some theorems on conformal transformations of Riemannian spaces*, *Proceedings of the Physico-Mathematical Society of Japan*, IIIs, vol. 18 (1936), pp. 572-578, and V. Modesitt, *Some singular properties of conformal transformations between Riemann spaces*, *American Journal of Mathematics*, vol. 60 (1938), pp. 325-336. Also see an abstract by the author entitled *Conformal transformations and the subspaces of a Riemann space*, *Bulletin of the American Mathematical Society*, abstract 43-9-328.

|| Throughout this paper except where otherwise stated Latin indices have the range 1, 2, ...,  $n$ . An index which appears twice in an expression is to be summed over the appropriate range unless the index appears in parentheses. A free index in a tensor equation assumes each value of its range.

It is assumed that these forms are not singular although they may be indefinite. We choose coordinate systems  $\{x^i\}$  and  $\{\bar{x}^i\}$  so that the conformal correspondence becomes  $\bar{x}^i = x^i$ . In these coordinate systems

$$(2.4) \quad \bar{g}_{ij} = e^{2\sigma} g_{ij}, \quad \bar{g}^{ij} = e^{-2\sigma} g^{ij},$$

where  $g^{ij}$  and  $\bar{g}^{ij}$  are the contravariant components of the metric tensors. If the Christoffel symbols of the second kind for  $V_n$  and  $\bar{V}_n$  are written  $\{k|ij\}$  and  $\{\bar{k}|\bar{i}\bar{j}\}$ , respectively, it follows from (2.4) that\*

$$(2.5) \quad \{\bar{k}|\bar{i}\bar{j}\} = \{k|ij\} + \delta_i^k \sigma_{,j} + \delta_j^k \sigma_{,i} - g_{ij} g^{km} \sigma_{,m}.$$

Let  $C$  be a curve in  $V_n$  and  $\bar{C}$  its image in  $\bar{V}_n$ . If the unit tangents to  $C$  and  $\bar{C}$  at corresponding points are denoted by  $\xi^i$  and  $\bar{\xi}^i$  and the principal normals by  $\mu^i$  and  $\bar{\mu}^i$ , it is an easy consequence of (2.4) and (2.5) that

$$(2.6) \quad \bar{\xi}^i = e^{-\sigma} \xi^i,$$

$$(2.7) \quad \bar{\mu}^i = e^{-2\sigma} [\mu^i - e \sigma_{,m} (g^{im} - e \xi^i \xi^m)],$$

where  $e$  is  $+1$  or  $-1$ , being determined by  $e = g_{ij} \xi^i \xi^j$ .

Since  $\bar{\mu}^i = 0$  for the geodesics of  $\bar{V}_n$ , the characteristic equation of a family of conformal geodesics in  $V_n$  is, according to (2.7),

$$(2.8) \quad \mu^i = e \sigma_{,m} (g^{im} - e \xi^i \xi^m).$$

In the derivation of this equation and throughout the paper we exclude those conformal geodesics which are tangent to null vectors. Now  $g^{im} - e \xi^i \xi^m$  is the projection tensor† for the vector space normal to  $\xi^i$ . Hence we have as a result of (2.8) the following theorem:

**THEOREM 2.1.** *The principal normal of any curve of a family of conformal geodesics passing through a common point in a non-null direction is, except for sign, the projection normal to this direction of a fixed vector; the sign is determined by the character of the tangent to the curve.*

This is the first characteristic property of conformal geodesics. Of course the fixed vector is the gradient  $\sigma_{,i}$ . As a consequence of Theorem 2.1, we have the following equivalent theorem:

**THEOREM 2.2.** *If the curves of a family of conformal geodesics which pass through a common point of  $V_n$  are projected orthogonally upon the tangent flat  $S_n$  at that point, the centers of curvature of the  $\infty^{n-1}$  projections at the common point*

\* The comma denotes covariant differentiation with respect to the  $x$ 's and the form (2.1), and the  $\delta_i^k$  are the Kronecker deltas.

† Duschek-Mayer, *Lehrbuch der Differentialgeometrie*, vol. 2, 1930, pp. 44-45.

will lie on a flat  $S_{n-1}$  orthogonal to the fixed vector of Theorem 2.1, and the  $\infty^{n-1}$  osculating circles of the projections will have a second point in common.

The proof is immediate. For the principal normals of the conformal geodesics in  $V_n$  are identical with the principal normals of their orthogonal projections in the flat  $S_n$ . From Theorem 2.1, it follows that the end points of the principal normals of the projections lie on a spherical  $S_{n-1}$  whose diameter is the length of the gradient  $\sigma_{,i}$ . By inversion, it is seen that the centers of curvature lie on a hyperplane of  $S_n$  normal to the direction of  $\sigma_{,i}$ . According to (2.8), the linear vector space determined by the tangent and principal normal of any curve of (2.8) contains  $g^{im}\sigma_{,m}$ . This proves the next theorem:

**THEOREM 2.3.** *The osculating geodesic surfaces of the curves of a family of conformal geodesics which pass through a common point form a bundle of surfaces; they all contain the fixed vector of Theorem 2.1.*

If we omit the condition that the fixed vector of Theorem 2.1 be a gradient, it follows easily that Theorem 2.1 is the characteristic property of all families of curves whose equations are of the form

$$(2.9) \quad \mu^i = e\tau_m(g^{im} - e\xi^i\xi^m),$$

where  $\tau_i$  is an arbitrary vector. The solutions of (2.9) have been called *velocity systems* because of their connection with motion in fields of force. A geometric definition of velocity systems is possible. For consider the Weyl geometry\* whose coefficients of connection  $\Gamma_{jk}^i$  are

$$\Gamma_{jk}^i = \{i|jk\} + \delta_j^i\tau_k + \delta_k^i\tau_j - g_{jk}g^{im}\tau_m.$$

Then if the points in this Weyl space and  $V_n$  which have the same coordinates correspond, it follows easily that the velocity system (2.9) consists of the images in  $V_n$  of the paths of this Weyl geometry. It is clear that Theorems 2.1, 2.2, and 2.3 hold for all velocity systems.

**3. Property two: hyperosculating geodesic circles.** Velocity systems are characterized by Theorem 2.1. It remains to distinguish geometrically the families of conformal geodesics among the totality of velocity systems. For this purpose, we shall consider the osculating geodesic circles of the curves.

Let  $C$  be a curve in  $V_n$ , and denote the unit tangent, and unit normals of orders 1, 2,  $\dots$ ,  $n-1$ , and the first, second,  $\dots$ ,  $(n-1)$ st curvatures of  $C$  by  $_{(1)}\xi^i$ , and  $_{(2)}\xi^i$ ,  $_{(3)}\xi^i$ ,  $\dots$ ,  $_{(n)}\xi^i$ , and  $k_1$ ,  $k_2$ ,  $\dots$ ,  $k_{n-1}$ , respectively. A geodesic

\* Such geometries were proposed by Weyl as the basis of a combined theory of gravitation and electro-dynamics. Cf. H. Weyl, *Space, Time, Matter*, English translation, 1921, pp. 125, 296, and L. P. Eisenhart, *Non-Riemannian Geometry*, American Mathematical Society Colloquium Publications, vol. 8, New York, 1927, pp. 81-82.

circle of  $V_n$  is defined as a curve whose first curvature in  $V_n$  is constant and whose second curvature is identically zero. The geodesic circle which is tangent to  $C$  and has the same first curvature as  $C$  at the point of tangency is called the *osculating geodesic circle* of  $C$  at the point.\* It follows from a fundamental existence theorem of differential equations that every curve for which  ${}_{(1)}\xi^i$ ,  ${}_{(2)}\xi^i$ , and  $k_1$  exist at a point has a unique osculating geodesic circle at that point.†

The Frenet equations of  $C$  are

$$(3.1) \quad \frac{\delta {}_{(m)}\xi^i}{\delta s} = -e_{m-1}k_{m-1}{}_{(m-1)}\xi^i + e_{m+1}k_m{}_{(m+1)}\xi^i,$$

$m = 1, 2, \dots, n-1; k_0 = 0,$

where

$$(3.2) \quad e_m = g_{ij}{}_{(m)}\xi^i{}_{(m)}\xi^j$$

and

$$(3.3) \quad g_{ij}{}_{(k)}\xi^i{}_{(m)}\xi^j = 0, \quad m \neq k,$$

and where  $\delta/\delta s$  denotes covariant differentiation with respect to arc length along  $C$  so that

$$\frac{\delta \lambda^i}{\delta s} = \frac{d\lambda^i}{ds} + \{i|jk\}\lambda^j{}_{(1)}\xi^k; \quad \frac{\delta \lambda_i}{\delta s} = \frac{d\lambda_i}{ds} - \{j|ik\}\lambda_j{}_{(1)}\xi^k.$$

The geodesic circle of  $C$  will have higher than second order contact; that is, it will hyperosculate  $C$  if and only if the values of  $\delta x^i/\delta s$ ,  $\delta^2 x^i/\delta s^2$ , and  $\delta^3 x^i/\delta s^3$  are the same at the point of tangency. Since  $\delta x^i/\delta s = {}_{(1)}\xi^i$ , the Frenet equations (3.1) show that these conditions are equivalent to‡

$$(3.4) \quad \frac{dk_1}{ds} = 0, \quad k_2 = 0$$

at the point of contact.

Since  $\xi^i = {}_{(1)}\xi^i$ ,  $e = e_1$ , and  $\mu^i = e_2 k_1 {}_{(2)}\xi^i$ , it follows from (2.9), (3.2), and

\* Lipke defines the osculating geodesic circle of  $C$  at a point as the curve of constant geodesic curvature which lies in the osculating geodesic  $V_2$  of  $C$  at the point and agrees with  $C$  in curvature and direction at the point. The results of §3 are valid for either definition of the osculating circle, but in later sections our present definition, based on the Frenet equations of a curve in  $V_n$  rather than in  $V_2$ , is more advantageous.

† Cf. Duschek-Mayer, loc. cit., pp. 62-64, for this theorem and the subsequent use of the Frenet equations. If  $k_p$  is identically zero on a curve, it is to be understood in (3.1) that  $k_{p+1} = \dots = k_{n-1} = 0$  and that  ${}_{(p+1)}\xi^i, \dots, {}_{(n)}\xi^i$  are any vectors which satisfy (3.2) and (3.3).

‡ If  $n=2$ , the second of these equations should be omitted. Similar deletions are to be understood in equations (3.5) and (3.6).

(3.3) that for any velocity system

$$(3.5) \quad \tau_i (2) \xi^i = e_1 k_1, \quad \tau_i (r) \xi^i = 0, \quad r > 2.$$

If we differentiate (3.5) covariantly with respect to  $s$  and use (3.1) and (3.5), we find

$$(3.6) \quad \begin{aligned} (\tau_{i,j} - \tau_j \tau_i) (2) \xi^i (1) \xi^j &= e_1 \frac{dk_1}{ds}, \\ (\tau_{i,j} - \tau_j \tau_i) (3) \xi^i (1) \xi^j &= e_1 e_2 k_1 k_2, \\ (\tau_{i,j} - \tau_j \tau_i) (1+r) \xi^i (1) \xi^j &= 0, \quad r > 2, \end{aligned}$$

since  $d\tau_i/ds = \tau_{i,j} (1) \xi^j$ .

According to (3.4) and (3.6), the directions  $(1) \xi^i$  at a fixed point in which velocity curves are hyperosculated by their osculating geodesic circles are given by

$$(3.7) \quad \tau_{ij} (s) \xi^i (1) \xi^j = 0, \quad s > 1,$$

where

$$(3.8) \quad \tau_{ij} = \tau_{i,j} - \tau_i \tau_j.$$

The tensor  $\tau_{ij}$  is symmetric when and only when  $\tau_i$  is a gradient  $\sigma_i$ . We call the directions in which hyperosculatation occurs the *H-directions* of the velocity system. If we write  $e_1 \rho_1 = \tau_{ij} (1) \xi^i (1) \xi^j$ , it follows from (3.2), (3.3), and (3.7) that

$$(3.9) \quad (\tau_{ij} - \rho_1 g_{ij}) (m) \xi^i (1) \xi^j = 0.$$

Since the vectors  $(m) \xi^i$  are independent,

$$(\tau_{ij} - \rho_1 g_{ij}) (1) \xi^i = 0,$$

so that  $\rho_1$  is a root of the determinant equation  $|\tau_{ij} - \rho g_{ij}| = 0$ , and  $(1) \xi^i$  is a principal direction determined by  $\tau_{ij}$ . In general there are  $n$  distinct principal directions. This proves the following theorem:

**THEOREM 3.1.** *The H-directions of any velocity system are identical with the principal directions determined by the tensor (3.8) which are not tangent to null vectors. The velocity system is a family of conformal geodesics if and only if this tensor is symmetric.*

If  $\tau_{ij}$  is a symmetric tensor and none of the corresponding principal directions are null vectors (as is always the case if (2.1) is definite), it follows from the known theory\* that there exist  $n$  mutually orthogonal non-null principal

\* L. P. Eisenhart, *Riemannian Geometry*, 1926, pp. 107-112.

directions. Conversely, let  ${}_{(p)}\lambda^i$  be  $n$  principal directions determined by  $\tau_{ij}$  such that  $g_{ij} {}_{(p)}\lambda^i {}_{(q)}\lambda^j = 0$  if  $p \neq q$ . Then  $\tau_{ij} {}_{(p)}\lambda^i {}_{(q)}\lambda^j = 0$ , ( $p \neq q$ ). Any two arbitrary vectors  $\alpha^i, \beta^i$  may be written as

$$\alpha^i = \sum_{t=1}^n a_t {}_{(t)}\lambda^i, \quad \beta^i = \sum_{t=1}^n b_t {}_{(t)}\lambda^i,$$

where  $a_t$  and  $b_t$  are constants so that

$$\tau_{ij} \alpha^i \beta^j = \sum_{t=1}^n a_t b_t \tau_{ij} {}_{(t)}\lambda^i {}_{(t)}\lambda^j = \tau_{ij} \beta^i \alpha^j.$$

Hence  $\tau_{ij}$  is a symmetric tensor so that  $\tau_i = \sigma_{,i}$ . This completes the characterization of conformal geodesics stated in the next theorem:

**THEOREM 3.2.** *A family of conformal geodesics in a  $V_n$  whose first fundamental form is definite admits an orthogonal ennuple of  $H$ -directions at each point. Conversely, if a velocity system in any  $V_n$  admits  $n$  mutually orthogonal  $H$ -directions at each point, it is a family of conformal geodesics.*

**4. The  $H$ -directions.** We consider the  $H$ -directions of a family of conformal geodesics in greater detail. If the conformal correspondence between  $V_n$  and  $\bar{V}_n$  is given by (2.3), in accordance with Theorem 3.1, the  $H$ -directions of the images in  $V_n$  of the geodesics of  $\bar{V}_n$  coincide with the non-null principal directions determined by the tensor

$$(4.1) \quad \sigma_{ij} = \sigma_{,ij} - \sigma_{,i}\sigma_{,j}.$$

The  $H$ -directions of the images in  $\bar{V}_n$  of the geodesics of  $V_n$  are similarly determined by the tensor\*

$$(4.2) \quad \bar{\sigma}_{ij} = (-\sigma)_{,ij} - (-\sigma)_{,i}(-\sigma)_{,j}.$$

It follows readily from (2.5), (4.1), and (4.2) that

$$(4.3) \quad \bar{\sigma}_{ij} = -\sigma_{ij} - \Delta_1 \sigma g_{ij},$$

where  $\Delta_1 \sigma = g^{ij} \sigma_{,i}\sigma_{,j}$ .

Now (4.3) is an equation of the form†

$$(4.4) \quad \nu_{ij} = \sum_{k=1}^m a_k {}_{(k)}\nu_{ij} + b g_{ij},$$

where the  ${}_{(k)}\nu_{ij}$  and  $\nu_{ij}$  are symmetric tensors of the second order and the  $a_k$

\* The semicolon here denotes covariant differentiation with respect to the  $x$ 's and the form (2.2).

† Here  $(k)$  denotes the tensor and  $ij$  the components.

and  $b$  are scalars. If  $\lambda^i$  is a common principal direction in  $V_n$  determined by each of the  ${}_{(k)}\nu_{ij}$ , quantities  $\rho_k$  exist such that

$$(4.5) \quad ({}_{(k)}\nu_{ij} - \rho_k g_{ij})\lambda^i = 0, \quad k = 1, 2, \dots, m.$$

It follows from (4.4) and (4.5) that  $(\nu_{ij} - \rho g_{ij})\lambda^i = 0$ , where  $\rho = \sum_{k=1}^m a_k \rho_k + b$ , so that  $\lambda^i$  is also a principal direction determined by  $\nu_{ij}$ .

This result shows that the principal directions in  $V_n$  determined by  $\sigma_{ij}$  and  $\bar{\sigma}_{ij}$  coincide. Furthermore, according to (2.4), the principal directions determined by any tensor in  $V_n$  and the tensor having the same components in  $\bar{V}_n$  correspond. Hence the  $H$ -directions in  $V_n$  and  $\bar{V}_n$  correspond by means of the conformal transformation. Since the  $H$ -directions in  $V_n$  and  $\bar{V}_n$  are conformally equivalent, the mapping determines a unique set of  $H$ -directions. We call these  $H$ -directions the *H-directions of the conformal transformation* (2.3).

If  $R_{hijk}$  and  $\bar{R}_{hijk}$  are the Riemann curvature tensors of  $V_n$  and  $\bar{V}_n$ , it follows from (2.5) by straightforward calculation\* that

$$(4.6) \quad e^{-2\sigma} \bar{R}_{hijk} = R_{hijk} + g_{hk}\sigma_{ij} + g_{ij}\sigma_{hk} - g_{hj}\sigma_{ik} - g_{ik}\sigma_{hj} \\ + (g_{hk}g_{ij} - g_{hj}g_{ik})\Delta_1\sigma.$$

By means of (2.4) and (4.6), we find

$$(4.7) \quad (n-2)\sigma_{ij} = \bar{R}_{ij} - R_{ij} - g_{ij}[\Delta_2\sigma + (n-2)\Delta_1\sigma],$$

where  $R_{ij}$  and  $\bar{R}_{ij}$  are the Ricci tensors of  $V_n$  and  $\bar{V}_n$ , respectively, and  $\Delta_2\sigma = g^{ij}\sigma_{,ij}$ . Thus, if  $n > 2$ , the  $H$ -directions of the conformal transformation are the non-null principal directions determined by  $\bar{R}_{ij} - R_{ij}$ . We state these results in the following theorem:

**THEOREM 4.1.** *Let  $V_n$  and  $\bar{V}_n$  be two conformal Riemann spaces. Then the  $H$ -directions of the images in  $V_n$  of the geodesics of  $\bar{V}_n$  and of the images in  $\bar{V}_n$  of the geodesics of  $V_n$  correspond under the mapping. If  $n > 2$ , these  $H$ -directions coincide with the principal directions determined by  $R_{ij} - \bar{R}_{ij}$  which are not tangent to null vectors.*

Since (4.7) is of the form (4.4), we conclude that if  $n > 2$  a principal direction determined by two of the tensors  $\sigma_{ij}$ ,  $R_{ij}$ ,  $\bar{R}_{ij}$  is also determined by the third. This proves the next theorem:

**THEOREM 4.2.** *Let  $V_n$  and  $\bar{V}_n$  be two conformal Riemann spaces of dimensionality  $n > 2$ . Then by the mapping of  $V_n$  on  $\bar{V}_n$ , a non-null Ricci principal direction of  $V_n$  corresponds to a Ricci principal direction of  $\bar{V}_n$  if and only if it is an  $H$ -direction of the conformal transformation.*

\* For example, cf. Eisenhart, loc. cit., pp. 89-90.

## II. PARTICULAR RIEMANN SPACES

5. **Conjugate conformal geodesics at a point.** Let  $\lambda^i$  be any unit vector at a point  $P$  of  $V_n$  which is not tangent to  $\sigma_{,i}$ . Then there exists a unique unit vector\*  $\lambda'^i$  in the linear vector space at  $P$  determined by  $\lambda^i$  and  $\sigma_{,i}$  such that

$$g_{ij}\lambda^i\lambda'^j = 0.$$

Let  $C$  and  $C'$  be the curves belonging to a family of conformal geodesics which are tangent to  $\lambda^i$  and  $\lambda'^i$ , respectively. Then  $C'$  is called the *conjugate conformal geodesic* of  $C$  at  $P$ . If  $\lambda^i\sigma_{,i} \neq 0$ , it is clear that this relationship is reciprocal, so that we may speak of  $C$  and  $C'$  as conjugate conformal geodesics. If  $\lambda^i\sigma_{,i} = 0$ , the conjugate conformal geodesic of  $C'$  is not defined.

According to Theorem 2.1, if  $\lambda^i\sigma_{,i} \neq 0$ , the unit first normals of  $C$  and  $C'$  at  $P$  are  $\lambda'^i$  and  $\lambda^i$ , respectively. Hence conjugate conformal geodesics have the same osculating geodesic surface. From (3.5), we have at  $P$

$$(5.1) \quad \sigma_{,i}\lambda'^i = ek, \quad \sigma_{,i}\lambda^i = e'k',$$

where  $k$  and  $k'$  are the first curvatures of  $C$  and  $C'$ , respectively, and

$$(5.2) \quad e = g_{ij}\lambda^i\lambda^j, \quad e' = g_{ij}\lambda'^i\lambda'^j.$$

If (2.8) is multiplied by  $\sigma_{,i}$  and summed for  $i$ , it follows from (5.1) that

$$(5.3) \quad ek'^2 + e'k^2 = \Delta_1\sigma.$$

If  $C$  is any conformal geodesic orthogonal to  $\sigma_{,i}$  at  $P$ , the equation corresponding to (5.3) is

$$(5.4) \quad e_\sigma k_\sigma^2 = \Delta_1\sigma,$$

where  $k_\sigma$  is the first curvature of  $C$  and where  $e_\sigma$  is +1 or -1 according as  $\Delta_1\sigma$  is positive or negative. If  $\Delta_1\sigma = 0$ ,  $k_\sigma = 0$ .

As  $P$  moves along  $C$ , we obtain by covariant differentiation of (5.1) with respect to  $s$ , after using (3.1), (3.5), (5.3), and (5.4), that

$$(5.5) \quad \sigma_{,ij}\lambda'^i\lambda^j = e \frac{dk}{ds},$$

$$(5.6) \quad \sigma_{,ij}\lambda^i\lambda'^j = e' \frac{dk'}{ds} - ee_\sigma k_\sigma^2,$$

where  $\sigma_{,ij}$  is defined by (4.1). Since  $\sigma_{,ij}$  is a symmetric tensor, if  $\lambda^i\sigma_{,i} \neq 0$  it follows from (5.5) that  $edk/ds = e'dk'/ds'$ , where  $s'$  denotes the arc length

\* An exception occurs only if  $\lambda^i\sigma_{,i} = 0$  and  $\Delta_1\sigma = 0$ . In this case,  $\lambda'^i$  is tangent to  $\sigma_{,i}$  and is a null vector.

of  $C'$ . If  $n=2$ , it can be shown that this last equation is equivalent to the second characteristic property of conformal geodesics.

**6. Conformal images of the geodesics of an  $S_n$  or  $E_n$ .** The geometric characterization of any family of conformal geodesics in  $V_n$  is derived in §§2 and 3. If these curves are the images of the geodesics of a space of constant curvature or an Einstein space, they will enjoy additional special properties which are obtained below. Let  ${}_{(\alpha)}\lambda^i$  and  ${}_{(\beta)}\lambda^i$  be any two unit vectors in  $V_n$  at a point  $P$  neither of which is tangent to  $\sigma_{,i}$ . We denote by  $C_\alpha$  and  $C_\beta$  the conformal geodesics tangent to these vectors and by  $C'_\alpha$  and  $C'_\beta$  the corresponding conjugate conformal geodesics.

According to (2.6) and (4.6),

$$(6.1) \quad e^{2\sigma}\bar{r}_{\alpha\beta} = r_{\alpha\beta} - e_\beta\sigma_{ik}{}_{(\beta)}\lambda^i{}_{(\beta)}\lambda^j - e_\alpha\sigma_{hj}{}_{(\alpha)}\lambda^h{}_{(\alpha)}\lambda^j - \Delta_1\sigma,$$

where  $r_{\alpha\beta}$  and  $\bar{r}_{\alpha\beta}$  are the Riemannian curvatures at  $P$  of  $V_n$  and  $\bar{V}_n$ , respectively, for the orientation determined by  ${}_{(\alpha)}\lambda^i$  and  ${}_{(\beta)}\lambda^i$ , and the  $e$ 's are defined in a manner analogous to (5.2). It follows from (5.4), (5.6), and (6.1) that

$$(6.2) \quad e^{2\sigma}\bar{r}_{\alpha\beta} = r_{\alpha\beta} - e_\alpha e'_\alpha \frac{dk'_\alpha}{ds_\alpha} - e_\beta e'_\beta \frac{dk'_\beta}{ds_\beta} + e_\sigma k_\sigma^2,$$

where  $k'_\alpha$  and  $k'_\beta$  are the first curvatures of  $C'_\alpha$  and  $C'_\beta$  and where  $s_\alpha$  and  $s_\beta$  are the arc lengths of  $C_\alpha$  and  $C_\beta$ , respectively.

If  $\bar{V}_n$  is a space of constant curvature  $K_0$ , the right-hand member of (6.2) does not depend upon the orientation determined by  ${}_{(\alpha)}\lambda^i$  and  ${}_{(\beta)}\lambda^i$  but is a scalar function in  $V_n$ . The algebraic sign of this function is constant and agrees with that of  $K_0$ . Conversely if the right-hand member of (6.2) is a scalar function, it follows from (6.2) that  $\bar{V}_n$  has the same Riemann curvature for every orientation at  $P$ . By Schur's theorem the curvature of  $\bar{V}_n$  is a constant  $K_0$ . The sign of  $K_0$  is determined by the scalar function. Since by a magnification two spaces of constant positive (or negative) may be mapped on each other so that their geodesics correspond, the precise value of  $K_0$  must be indeterminate. We state these results in the following theorem:

**THEOREM 6.1.** *The necessary and sufficient condition that a family of conformal geodesics in  $V_n$  be the images of the geodesics of an  $S_n$  is that*

$$r_{\alpha\beta} - e_\alpha e'_\alpha \frac{dk'_\alpha}{ds_\alpha} - e_\beta e'_\beta \frac{dk'_\beta}{ds_\beta} + e_\sigma k_\sigma^2$$

*be a point function in  $V_n$ . The  $S_n$  has positive, zero, or negative Riemann curvature according as this function is greater than, equal to, or less than zero.*

According to (2.6) and (4.7),

$$(6.3) \quad e_a(n-2)\sigma_{ij}(\alpha)\lambda^i(\alpha)\lambda^j = e^{2\sigma}\bar{\gamma}_a - \gamma_a - \Delta_2\sigma - (n-2)\Delta_1\sigma,$$

where  $\gamma_a$  and  $\bar{\gamma}_a$  are the Ricci or mean curvatures for the direction  $(\alpha)\lambda^i$  of  $V_n$  and  $\bar{V}_n$ , respectively. It follows from (5.6) and (6.3) that

$$(6.4) \quad (n-2)e_a e'_a \frac{dk'_a}{ds_a} + \gamma_a = (n-2)e_a k_a^2 + e^{2\sigma}\bar{\gamma}_a - \Delta_2\sigma - (n-2)\Delta_1\sigma.$$

If  $\bar{V}_n$ , ( $n > 2$ ), is an Einstein space,  $\bar{\gamma}_a$  is a constant. Hence the left-hand member of (6.4) is a scalar. The converse is also true. This proves the next theorem:

**THEOREM 6.2.** *The necessary and sufficient condition that a family of conformal geodesics in a  $V_n$  of dimensionality  $n > 2$  be the images of the geodesics of an  $E_n$  is that*

$$(n-2)e_a e'_a \frac{dk'_a}{ds_a} + \gamma_a$$

*be a point function in  $V_n$ .*

In addition to this characteristic property, we easily obtain further necessary properties of the conformal images of the geodesics of an Einstein space.

**THEOREM 6.3.** *If a family of conformal geodesics in a  $V_n$  of dimensionality  $n > 2$  are the images of the geodesics of an  $E_n$ , the  $H$ -directions of the family coincide with the principal Ricci directions of  $V_n$  which are not tangent to null vectors.*

This is an immediate consequence of (2.4) and Theorem 4.1. For  $\bar{V}_n$  is an Einstein space if and only if

$$(6.5) \quad \bar{R}_{ij} = a\bar{g}_{ij},$$

where  $a$  is a constant.

**THEOREM 6.4.** *Let  $E_n$  and  $\bar{E}_n$  be conformal Einstein spaces of dimensionality  $n > 2$ . Then the conformal images in  $E_n$  of the geodesics of  $\bar{E}_n$  as well as the images in  $\bar{E}_n$  of the geodesics of  $E_n$  are geodesic circles.*

It has been shown by Brinkmann\* that a large class of Einstein spaces exist which are conformal to Einstein spaces. The above theorem applies to these spaces. The proof of the theorem follows. By definition of  $E_n$ ,

$$(6.6) \quad R_{ij} = b g_{ij},$$

\* H. W. Brinkmann, *Einstein spaces which are mapped conformally on each other*, Mathematische Annalen, vol. 94 (1925), pp. 119-145.

where  $b$  is a constant. It follows from (2.4), (4.7), (6.5), and (6.6) that

$$\sigma_{ij} = \phi g_{ij},$$

where  $\phi$  is a scalar. Hence every direction is an  $H$ -direction so that, as follows from (3.4),  $dk_1/ds=0$ ,  $k_2=0$  for each conformal geodesic. This proves Theorem 6.4. According to Theorem 4.1, the conformal geodesics of  $V_n$  and  $\bar{V}_n$ , ( $n>2$ ), will be geodesic circles if and only if the mapping is such that

$$\bar{R}_{ij} - R_{ij} = \psi g_{ij},$$

where  $\psi$  is a scalar function.

**7. Conformal geodesics in an  $S_n$ .** The geometric property of conformal geodesics in a  $V_n$  stated in Theorem 2.2 is not intrinsic since it depends upon the tangent flat  $S_n$ . The only exception arises when  $V_n$  is itself a flat space. In this case, Theorem 2.2 becomes: The centers of curvature of the curves of a family of conformal geodesics which pass through a common point lie on a flat  $S_{n-1}$  orthogonal to the direction of  $\sigma_{,i}$ . In what follows, we show that the spaces of constant curvature enjoy an analogous property. The results apply without modification to all velocity systems.

We begin by generalizing the notion of center of curvature to apply to a curve  $C$  in  $V_n$ . Let  $V_2$  be the osculating geodesic surface of  $C$  at  $P$ , and let  $C''$  be any curve in  $V_2$  which at  $P$  has the same tangent and principal normal\* as  $C$ . The limiting first point of intersection (when it exists) of the geodesics of  $V_2$  normal to  $C''$  at  $P$  and at a nearby point  $Q$  as  $Q$  approaches  $P$  is called the *center of curvature of  $C$  in  $V_n$  at the point  $P$* . From the viewpoint of the calculus of variations, the center of curvature of  $C$  is the focal point of  $C''$  on the geodesic normal to  $C''$  at  $P$ .† Since the focal point depends only on the first curvature of  $C''$  at  $P$ , if it exists it is uniquely determined by  $C$ .

In accordance with Theorem 2.3, the osculating geodesic surfaces of the curves of a family of conformal geodesics  $\{C\}$  which pass through a common point  $P$  all contain the gradient  $\sigma_{,i}$ . This means that a geodesic surface  $V_2$  at  $P$  osculates  $\infty^1$  conformal geodesics passing through  $P$ . The locus of the centers of curvature in  $V_n$  of these conformal geodesics is, in general, a curve in  $V_2$ . In what follows, we prove the following theorem:

**THEOREM 7.1.** *Let  $\{C\}$  be any family of conformal geodesics in a  $V_n$  whose first fundamental form is positive definite. Then the locus of the centers of curva-*

\* Since  $V_2$  is geodesic at  $P$ ,  $C''$  has the same first curvature at  $P$  when considered as a curve in  $V_2$  or  $V_n$ .

† O. Bolza, *Vorlesungen über Variationsrechnung*, 1909, p. 323, and M. Morse, *The Calculus of Variations in the Large*, American Mathematical Society Colloquium Publications, vol. 18, New York, 1934, pp. 51-55.

ture in  $V_n$  (if they exist) of the  $\infty^1$  curves of  $\{C\}$  which pass through a common point osculating the same geodesic surface  $V_2$  at that point is a geodesic of  $V_2$  if and only if  $V_2$  is an  $S_2$ .

The existence of the center of curvature of a curve in  $S_n$  is discussed later. If we use the geodesic polar coordinates with center at  $P$ , the first fundamental form of  $V_2$  is

$$(7.1) \quad ds^2 = dr^2 + G(r, \theta)d\theta^2,$$

where

$$(7.2) \quad (G(0, \theta))^{1/2} = 0, \quad \frac{\partial(G(0, \theta))^{1/2}}{\partial r} = 1.$$

Now the center of curvature in  $V_n$  of a curve osculating  $V_2$  at  $P$  depends only on its direction and first curvature  $k_1$  at  $P$ . If  $(r, \theta)$  are the coordinates of the center of curvature, it follows that

$$(7.3) \quad k_1 = f(r, \theta).$$

The function  $f(r, \theta)$  is completely determined by the surface  $V_2$ . Indeed, it may be shown by the methods of the calculus of variations that

$$(7.4) \quad f(r, \theta) = - \int \frac{dr}{G(r, \theta)}, \quad \lim_{r \rightarrow 0} \left[ \int \frac{dr}{G(r, \theta)} + \frac{1}{r} \right] = 0.$$

This relation is not used in the present proof.

According to (3.5), the first curvatures of the  $\infty^1$  curves of  $\{C\}$  osculating  $V_2$  at  $P$  obey an equation of the form

$$(7.5) \quad k_1 = a \sin(\theta + b),$$

where  $a$  and  $b$  are constants which depend upon the particular family  $\{C\}$  and the point  $P$ .<sup>\*</sup> From (7.3) and (7.5), the locus of the centers of curvature of the  $\infty^1$  conformal geodesics is

$$(7.6) \quad f(r, \theta) = a \sin(\theta + b).$$

By hypothesis this locus is a geodesic of  $V_2$  for every value of  $a$  and  $b$ . Differentiating (7.6), we find that the curves (7.6) are the solutions of

$$(7.7) \quad \frac{d^2 r}{d\theta^2} + \frac{f_{rr}}{f_r} \left( \frac{dr}{d\theta} \right)^2 + \frac{2f_{r\theta}}{f_r} \left( \frac{dr}{d\theta} \right) + \frac{f_{\theta\theta} + f}{f_r} = 0.$$

But (7.6) must also satisfy the differential equation for the geodesics of  $V_2$ :

<sup>\*</sup> Indeed, for the family  $\{C\}$  associated with the conformal mapping (2.3),  $a = (\Delta_1 \sigma)^{1/2}$  and  $b$  is chosen so that the vector  $\sigma_{,4}$  is tangent to  $\theta = -b$ .

$$(7.8) \quad \frac{d^2 r}{d\theta^2} - \frac{G_r}{G} \left( \frac{dr}{d\theta} \right)^2 - \frac{G_\theta}{2G} \left( \frac{dr}{d\theta} \right) - \frac{G_r}{2} = 0.$$

It follows from (7.7) and (7.8) that each curve (7.6) will be a geodesic if and only if

$$(7.9) \quad \frac{f_{rr}}{f_r} + \frac{G_r}{G} = 0, \quad \frac{2f_{\theta r}}{f_r} + \frac{G_\theta}{2G} = 0, \quad \frac{f_{\theta\theta} + f}{f_r} + \frac{G_r}{2} = 0.$$

A little calculation shows that the solution of (7.2) and (7.9) is (7.4) and one of the following equations:

$$(7.10) \quad G(r, \theta) = \frac{1}{c^2} \sin^2 cr, \quad G(r, \theta) = r^2, \quad G(r, \theta) = \frac{1}{c^2} \sinh^2 cr.$$

Since the Gaussian curvature of  $V_2$  is equal to  $-(G^{1/2})_{,rr}/G^{1/2}$ , it follows from (7.10) that  $V_2$  is a surface of constant curvature  $c^2$ , 0, or  $-c^2$ , respectively. This completes the proof.

If  $V_n$  is an  $S_n$ , it follows easily from (7.4), (7.5), and (7.10) that the locus of centers of curvature in any geodesic  $S_2$  is\*

$$(7.11) \quad c \cot cr = (\Delta_1 \sigma)^{1/2} \sin \theta, \quad \frac{1}{r} = (\Delta_1 \sigma)^{1/2} \sin \theta, \quad c \coth cr = (\Delta_1 \sigma)^{1/2} \sin \theta,$$

according as  $S_n$  has Riemann curvature  $c^2$ , 0, or  $-c^2$ , respectively. Hence the locus always exists in an  $S_n$  of positive or zero curvature and exists in an  $S_n$  of negative Riemann curvature  $-c^2$  if and only if  $\Delta_1 \sigma \sin^2 \theta > c^2$ .

In an  $S_n$ , every geodesic  $V_2$  is a totally geodesic  $S_2$ . Therefore, the locus of the centers of curvature in any  $S_2$  of the appropriate curves passing through  $P$  of a family of conformal geodesics  $\{C\}$  is a geodesic of  $S_n$ . This geodesic is easily shown to be orthogonal to the geodesic of  $S_n$  which is tangent to  $\sigma_{,i}$  at  $P$ . Furthermore, the point at which the two geodesics intersect orthogonally does not depend upon the particular osculating geodesic surface at  $P$ . Hence the totality of geodesics in  $S_n$  which are the loci associated with the family  $\{C\}$  lie on a totally geodesic  $S_{n-1}$  orthogonal to the geodesic which is tangent to  $\sigma_{,i}$ . We state this result in the following theorem:

**THEOREM 7.2.** *Let  $\{C\}$  be a family of conformal geodesics in an  $S_n$  whose first fundamental form is positive definite. Then the centers of curvature in  $S_n$  (if they exist) of the curves of  $\{C\}$  which pass through a common point lie on a totally geodesic  $S_{n-1}$  orthogonal to the geodesic through the point which is tangent to the fixed vector of Theorem 2.1.*

\* The direction of the gradient  $\sigma_{,i}$  is  $\theta=0$ .

III. SUBSPACES OF  $V_n$ 

8. **Conformal geodesics in subspaces of  $V_n$ .** The conformal transformation (2.3) induces a conformal mapping of the respective subspaces of  $V_n$  and  $\bar{V}_n$  upon each other. If  $V_p$  and  $\bar{V}_p$ , ( $1 < p < n$ ), are two such conformal subspaces, the images of the geodesics of  $\bar{V}_p$  in  $V_p$  are conformal geodesics of  $V_p$  and therefore enjoy the properties of conformal geodesics derived in the previous sections. We now consider the additional relationships which exist between the families of conformal geodesics of  $V_n$  and of its subspaces which are the conformal images of the geodesics of  $\bar{V}_n$  and its respective subspaces under the transformation (2.3). We refer to these conformal geodesics of  $V_n$  and of its subspaces as *corresponding* families of conformal geodesics. Any one such family is said to *correspond* to the transformation (2.3). It is clear that if a family of conformal geodesics of  $V_n$  is given, then the conformal transformation is determined except for a magnification. Hence all the corresponding families of conformal geodesics in the subspaces of  $V_n$  are uniquely determined by the given conformal geodesics of  $V_n$ .

The equations of the imbedding of a  $V_p$  in  $V_n$  are\*  $x^i = x^i(y^\alpha)$ . If the first fundamental form of  $V_n$  is (2.1), the corresponding form of  $V_p$  is

$$(8.1) \quad ds^2 = h_{\alpha\beta} dy^\alpha dy^\beta,$$

where†

$$(8.2) \quad h_{\alpha\beta} = g_{ij} x^i_{,\alpha} x^j_{,\beta}.$$

If  $\zeta^\alpha$  is a unit vector in  $V_p$ , the principal normal  $\mu^i$  of the conformal geodesic of  $V_n$  tangent to  $\zeta^\alpha$  is given by (2.8), where  $\xi^i$  are the components in the  $x$ 's of  $\zeta^\alpha$  so that

$$(8.3) \quad \xi^i = \zeta^\alpha x^i_{,\alpha}.$$

The principal normal  $\nu^\alpha$  of the corresponding conformal geodesic of  $V_p$  tangent to  $\zeta^\alpha$  is

$$\nu^\alpha = e\sigma_{,\beta}(h^{\alpha\beta} - e\zeta^\alpha\zeta^\beta),$$

where  $\sigma_{,\beta}$  given by

$$(8.4) \quad \sigma_{,\beta} = \sigma_{,m} x^m_{,\beta}$$

is the projection of  $\sigma_{,i}$  in the tangent vector space of  $V_p$ . If  $\eta^i$  are the components of  $\nu^\alpha$  in the  $x$ 's,  $\eta^i = \nu^\alpha x^i_{,\alpha}$ , so that

\* In this section, the range of the Greek letters is 1, 2, ...,  $p$  unless otherwise stated.

† The comma followed by a Greek letter denotes covariant differentiation with respect to the form (8.1) and the  $y$ 's.

$$(8.5) \quad \eta^i = e\sigma_{,m}x_{,\beta}^m x_{,\alpha}^i (h^{\alpha\beta} - e\xi^\alpha \xi^\beta).$$

It is easy to show from the canonical representation for  $h^{\alpha\beta}$  and  $g^{ij}$  that

$$(8.6) \quad h^{\alpha\beta} x_{,\alpha}^i x_{,\beta}^m = g^{im} - \sum_{\tau=p+1}^n e_\tau ({}_r\lambda^i {}_r\lambda^m),$$

where the  ${}_r\lambda^i$  are  $n-p$  mutually orthogonal unit normals to  $V_p$  and  $e_\tau = g^{ij} {}_r\lambda^i {}_r\lambda^j$ . It follows from (2.8), (8.3), (8.5), and (8.6) that\*

$$(8.7) \quad \eta^i = \mu^i - e\sigma_{,m} \sum_\tau e_\tau {}_r\lambda^m {}_r\lambda^i.$$

It is clear that the last expression in (8.7) is the signed projection of  $\sigma_{,m}$  in the normal vector space of  $V_p$ . As an immediate consequence of (8.7), it is seen that *if two  $V_p$ 's are tangent at a point, the principal normals of their corresponding conformal geodesics which pass through the point in the same direction are equal.*

While the principal normal is thus determined by the tangent vector space of  $V_p$ , the  $H$ -directions also depend upon the tensors  ${}_r\Omega_{\alpha\beta}$ . These tensors are introduced in the equations†

$$(8.8) \quad x_{,\alpha\beta}^i = -\{i|jk\} x_{,\alpha}^j x_{,\beta}^k + \sum_\tau e_\tau ({}_r\Omega_{\alpha\beta} {}_r\lambda^i)$$

and are used to construct the second fundamental form of  $V_p$ . According to Theorem 3.1, the  $H$ -directions of the corresponding family of conformal geodesics of  $V_p$  are the non-null principal directions determined by the tensor  $\sigma_{\alpha\beta}$  where

$$(8.9) \quad \sigma_{\alpha\beta} = \sigma_{,\alpha\beta} - \sigma_{,\alpha}\sigma_{,\beta}.$$

By straightforward calculation, we find from (4.1), (8.4), (8.8), and (8.9) that

$$(8.10) \quad \sigma_{\alpha\beta} = \sigma_{ij} x_{,\alpha}^i x_{,\beta}^j + \sum_\tau e_\tau ({}_r\lambda^i \sigma_{,i} {}_r\Omega_{\alpha\beta}).$$

If  $\xi^i$  is an  $H$ -direction at a point  $P$  of a family of conformal geodesics of  $V_n$ , in accordance with Theorem 3.1,

$$(8.11) \quad (\sigma_{ij} - \rho g_{ij})\xi^i = 0,$$

where  $\rho = \sigma_{ij}\xi^i\xi^j$  and  $e = g_{ij}\xi^i\xi^j$ . For any  $V_p$  which contains  $\xi^i$ , the components  $\xi^\alpha$  in the  $y$ 's of  $\xi^i$  satisfy (8.3). We multiply (8.11) by  $x_{,\beta}^j$  and sum the

\* This equation may also be obtained directly from Theorem 2.1.

† Eisenhart, loc. cit., p. 160.

resulting equation for  $j$ . After using (8.2) and (8.3), this equation becomes

$$(\sigma_{ij}x_{,a}^ix_{,b}^j - \rho h_{ab})\xi^a = 0.$$

Hence  $\xi^a$  is a principal direction determined by the tensor  $\sigma_{ij}x_{,a}^ix_{,b}^j$ . But if  $V_p$  is geodesic or umbilical at a point  $P$ , it follows from (8.10) that the  $H$ -directions at  $P$  of the corresponding family of conformal geodesics of  $V_p$  are determined by  $\sigma_{ij}x_{,a}^ix_{,b}^j$ . This discussion proves the following theorem:

**THEOREM 8.1.** *If at a point an  $H$ -direction of a family of conformal geodesics of  $V_n$  is tangent to a subspace  $V_p$  which is geodesic or umbilical at the point, then this direction is also an  $H$ -direction of the corresponding family of conformal geodesics of  $V_p$ .*

We now suppose  $p = n - 1$ , and write (8.10) as

$$(8.12) \quad \sigma_{ab} = \sigma_{ij}x_{,a}^ix_{,b}^j + e_1(1)\lambda^i\sigma_{,i}\Omega_{ab}.$$

As was shown in the paragraph below (4.3), if  $\xi^i$  is a principal direction determined by two of the tensors  $\sigma_{ab}$ ,  $\sigma_{ij}x_{,a}^ix_{,b}^j$ ,  $\Omega_{ab}$ , it is also determined by the third provided  $(1)\lambda^i\sigma_{,i} \neq 0$ . But  $\Omega_{ab}$  determines the directions of the lines of curvature of  $V_{n-1}$ , and, as was shown above,  $\sigma_{ij}x_{,a}^ix_{,b}^j$  determines the  $H$ -directions of the geodesic  $V_{n-1}$  which has the same orientation as  $V_{n-1}$ . This proves the next theorem:

**THEOREM 8.2.** *Let  $V_{n-1}$  be a hypersurface of  $V_n$  which at a point does not contain the fixed vector of Theorem 2.1. Then if, at this point, a vector is a member of two of the following sets, it is also a member of the third set:*

- (1) *the tangents of the lines of curvature of  $V_{n-1}$ ,*
- (2) *the  $H$ -directions of a family of conformal geodesics of  $V_{n-1}$ ,*
- (3) *the  $H$ -directions of the corresponding family of conformal geodesics of the tangent geodesic  $V_{n-1}$ .*

As a consequence of Theorem 8.1 and Theorem 8.2, we note that an  $H$ -direction of a family of conformal geodesics of  $V_n$  is also an  $H$ -direction of the corresponding family of conformal geodesics of a hypersurface  $V_{n-1}$  with  $(1)\lambda^i\sigma_{,i} \neq 0$  if and only if the direction is tangent to a line of curvature of  $V_{n-1}$ .

Now the corresponding family of conformal geodesics of a hypersurface  $\sigma = \text{const.}$  is simply the totality of geodesics of the hypersurface; so the  $H$ -directions of this family are completely indeterminate. It follows from the statement italicized above that an  $H$ -direction of a family of conformal geodesics of  $V_n$  which lies in a hypersurface  $\sigma = \text{const.}$  with  $\Delta_1\sigma \neq 0$  is tangent to a line of curvature of the hypersurface.

We now consider two hypersurfaces  $V_{n-1}$  and  $V_{n-1}^*$  which are tangent at a point  $P$  of  $V_n$ . If their equations of imbedding in  $V_n$  are  $x^i = x^i(y^a)$  and  $x^i = x^{*i}(y^a)$ , it follows that at  $P$ ,  $\partial x^i/\partial y^a$  and  $\partial x^{*i}/\partial y^a$  span the same tangent vector space. Hence we may choose the coordinate directions  $y^a$  for  $V_{n-1}$  and  $V_{n-1}^*$  as mutually tangent at  $P$  so that at this point

$$(8.13) \quad \frac{\partial x^i}{\partial y^a} = \frac{\partial x^{*i}}{\partial y^a}.$$

Now the tensor  $\sigma_{ab}$  for  $V_{n-1}$  is given by (8.12), and the corresponding tensor  $\sigma_{ab}^*$  for  $V_{n-1}^*$  by

$$(8.14) \quad \sigma_{ab}^* = \sigma_{ij} x_{,a}^{*i} x_{,b}^{*j} + e_1 (1) \lambda^{*i} \sigma_{,i} \Omega_{ab}^*,$$

where the notation is analogous to that used in (8.12) and refers to  $V_{n-1}$ . Since at  $P$ ,  $(1)\lambda^i = (1)\lambda^{*i}$ , it follows from (8.12), (8.13), and (8.14) that at the point of contact,

$$(8.15) \quad \sigma_{ab} - \sigma_{ab}^* = e_1 (1) \lambda^i (\Omega_{ab} - \Omega_{ab}^*).$$

If  $(1)\lambda^i \sigma_{,i} = 0$ , it follows that  $\sigma_{ab} = \sigma_{ab}^*$ . More generally, if two  $V_n$ 's are tangent at  $P$  and contain the gradient  $\sigma_{,i}$  at  $P$ , the  $H$ -directions of the corresponding families of conformal geodesics coincide at this point.

If  $(1)\lambda^i \sigma_{,i} \neq 0$  and  $\zeta^a$  denotes a unit vector of  $V_{n-1}$  and  $V_{n-1}^*$  at  $P$ , we obtain

$$(8.16) \quad \sigma_{ab} \zeta^a \zeta^b - \sigma_{ab}^* \zeta^a \zeta^b = e_1 (1) \lambda^i \sigma_{,i} (\Omega_{ab} \zeta^a \zeta^b - \Omega_{ab}^* \zeta^a \zeta^b).$$

According to (5.6),

$$(8.17) \quad \sigma_{ab} \zeta^a \zeta^b = e' \frac{dk'}{ds} - ee_\sigma k_\sigma^2, \quad \sigma_{ab}^* \zeta^a \zeta^b = e' \frac{dk^{*'}}{ds^*} - ee_\sigma k_\sigma^{*2},$$

where the notation is analogous to that of (5.6) and refers to the corresponding families of conformal geodesics of  $V_{n-1}$  and  $V_{n-1}^*$ . The remarks following (8.7) show that

$$(8.18) \quad k_\sigma = k_\sigma^*.$$

Of course

$$(8.19) \quad \Omega_{ab} \zeta^a \zeta^b = eK, \quad \Omega_{ab}^* \zeta^a \zeta^b = eK^*,$$

where  $K$  and  $K^*$  denote the normal curvatures of  $V_{n-1}$  and  $V_{n-1}^*$ , respectively, for the direction  $\zeta^a$ . It follows from (8.16), (8.17), (8.18), and (8.19) that†

† A similar equation may be obtained for a single  $V_{n-1}$  by using (8.12) instead of (8.15) in the above derivation.

$$(8.20) \quad \frac{dk'}{ds} - \frac{dk^{*'}}{ds^{*}} = e_1 e e'_{(1)} \lambda^i \sigma_{,i} (K - K^{*}).$$

Hence the difference of normal curvatures for the same direction on two tangent hypersurfaces is expressible in terms of the curvatures of the corresponding conformal geodesics of these hypersurfaces. It also follows from (8.20) that the expression

$$\left( \frac{dk'}{ds} - \frac{dk^{*'}}{ds^{*}} \right) \frac{1}{(1) \lambda^i \sigma_{,i}}$$

does not depend upon the conformal mapping (2.3) of  $V_n$  upon  $\bar{V}_n$ ; that is, it is invariant for any pair of corresponding families of conformal geodesics.

As an immediate consequence of (8.15) and the remarks below (4.3), we have the following theorem:

**THEOREM 8.3.** *Let  $V_{n-1}$  and  $V_{n-1}^{*}$  be tangent at a point where they do not contain the fixed vector of Theorem 2.1. Then if, at this point, a vector is a member of three of the following sets, it is also a member of the fourth set:*

- (1) *the tangents of the lines of curvature of  $V_{n-1}$ ,*
- (2) *the tangents of the lines of curvature of  $V_{n-1}^{*}$ ,*
- (3) *the  $H$ -directions of a family of conformal geodesics of  $V_{n-1}$ ,*
- (4) *the  $H$ -directions of the corresponding family of conformal geodesics of  $V_{n-1}^{*}$ .*

If the difference of the normal curvatures at  $P$  of  $V_{n-1}$  and  $V_{n-1}^{*}$  for the same direction is constant as the direction changes, it follows that  $\Omega_{\alpha\beta} = \Omega_{\alpha\beta}^{*} + a h_{\alpha\beta}$ , where  $a$  is a constant. In this case, according to (8.15), the  $H$ -directions for any corresponding families of conformal geodesics in  $V_{n-1}$  and  $V_{n-1}^{*}$  coincide at  $P$ .

A conformal transformation of  $V_n$  for which

$$(8.21) \quad \sigma_{ij} = \phi g_{ij}$$

has a particularly simple character. As noted in §6, it is only in this case that the corresponding conformal geodesics of  $V_n$  are geodesic circles. We investigate the induced conformal transformations of the hypersurfaces of  $V_n$ . From (8.2), (8.12), and (8.21),

$$(8.22) \quad \sigma_{\alpha\beta} = \phi h_{\alpha\beta} + e_1 (1) \lambda^i \sigma_{,i} \Omega_{\alpha\beta}$$

for any  $V_{n-1}$  in  $V_n$ . Hence if  $(1) \lambda^i \sigma_{,i} = 0$  at a point, every direction is an  $H$ -direction of the corresponding conformal geodesics of  $V_{n-1}$  at this point. If  $(1) \lambda^i \sigma_{,i} \neq 0$ , it follows from (8.22) that the non-null tangents to the lines of

curvature of  $V_{n-1}$  and the  $H$ -directions of the corresponding conformal geodesics of  $V_{n-1}$  coincide.

Conversely, suppose the direction of each line of curvature of any  $V_{n-1}$  in  $V_n$  is an  $H$ -direction of a family of conformal geodesics of  $V_{n-1}$  if the direction is not tangent to a null vector, and suppose that all of these families of conformal geodesics correspond to the same conformal mapping of  $V_n$ . Now it is easy to show\* that a  $V_{n-1}$  in  $V_n$  exists which contains an arbitrary point  $P$  of  $V_n$  and is such that the lines of curvature of  $V_{n-1}$  are tangent to an arbitrary enuple of non-null directions at  $P$ . Furthermore, we may choose coordinates  $y^\alpha$  in the  $V_{n-1}$  so that the tangents  $x_{,\alpha}^i$  ( $\alpha$  constant) to the coordinate lines are also tangent to the lines of curvature at  $P$ . In this coordinate system

$$\sigma_{\alpha\beta} = 0, \quad \Omega_{\alpha\beta} = 0, \quad \alpha \neq \beta.$$

It follows from (8.12) that  $\sigma_{ij}x_{,\alpha}^i x_{,\beta}^j = 0$ , ( $\alpha \neq \beta$ ). Since  $x_{,\alpha}^i$  and  $x_{,\beta}^j$  are arbitrary orthogonal vectors in  $V_n$ , the last equation shows that  $\sigma_{ij} = \phi g_{ij}$ . This proves the next theorem:

**THEOREM 8.4.** *Let  $V_n$  be conformal to  $\bar{V}_n$  so that the images of the geodesics of  $\bar{V}_n$  are geodesic circles in  $V_n$ . Then and only then the non-null tangents to the lines of curvature of any  $V_{n-1}$  in  $V_n$  are  $H$ -directions for the corresponding family of conformal geodesics in  $V_{n-1}$ .*

It is easy to see that non-trivial conformal transformations exist for which (8.21) holds. As noted in §6, the conformal mapping of any two Einstein spaces of dimensionality  $n > 2$  gives rise to an equation of the form (8.21). We discuss this topic further in §12.

As a consequence of Theorem 6.4 and the remarks following (8.22), we have the following theorem which may be illustrated by non-trivial examples:

**THEOREM 8.5.** *Let  $E_n$  and  $\bar{E}_n$  be conformal Einstein spaces of dimensionality  $n > 3$ , and let  $E_{n-1}$  and  $\bar{E}_{n-1}$  be Einstein hypersurfaces which correspond by the mapping and which do not contain the fixed vector of Theorem 2.1. Then  $E_{n-1}$  and  $\bar{E}_{n-1}$  have indeterminate lines of curvature.*

**9. The hypersurfaces  $\sigma = \text{const.}$**  For the conformal transformation (2.3), the hypersurfaces  $\sigma = \text{const.}$  play a special role. The mapping of these hypersurfaces in  $V_n$  and  $\bar{V}_n$ , respectively, upon each other is simply a change in scale. We investigate the conditions under which the normal to  $\sigma = \text{const.}$  may be an  $H$ -direction of the corresponding family of conformal geodesics of  $V_n$ . We obtain the following results:

\* For a proof of this statement, cf. A. Fialkow, *The Riemannian curvature of a hypersurface*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 256-257.

**THEOREM 9.1.** *The tangent at a point of  $V_n$  to a curve of the congruence normal to the hypersurfaces  $\sigma = \text{const.}$  is an  $H$ -direction of the corresponding family of conformal geodesics of  $V_n$  if and only if the curve is not tangent to a null vector and has zero first curvature at the point.*

The proof follows. According to Theorem 3.1 and the hypothesis that  $g^{im}\sigma_{,m}$  is an  $H$ -direction at a point  $P$  of  $V_n$ ,  $g^{im}\sigma_{,m}$  is a non-null principal direction determined by the tensor (4.1) at  $P$ . Hence, for a suitable  $\rho$ ,

$$(9.1) \quad (\sigma_{,ij} - \sigma_{,i}\sigma_{,j} - \rho g_{ij})\sigma^{,i} = 0$$

at this point, where  $\sigma^{,i} = g^{im}\sigma_{,m}$ . Let\*  ${}_{(p)}\lambda^i$  be  $n-1$  mutually orthogonal congruences of vectors in  $V_n$  such that

$$(9.2) \quad \sigma_{,i} {}_{(p)}\lambda^i = 0.$$

Differentiating (9.2) covariantly, we obtain

$$\sigma_{,ij} {}_{(p)}\lambda^j + \sigma_{,j} {}_{(p)}\lambda^j_{;i} = 0.$$

Hence

$$(9.3) \quad -\sigma_{,ij} {}_{(p)}\lambda^i \sigma^{,j} = {}_{(p)}\lambda^i_{;j} \sigma^{,j} \sigma^{,i}.$$

But from (9.1) and (9.2),  $\sigma_{,ij} {}_{(p)}\lambda^i \sigma^{,j} = 0$ . If we substitute this value in (9.3),

$$(9.4) \quad {}_{(p)}\lambda^i_{;j} {}_{(n)}\lambda^j {}_{(n)}\lambda^i = 0,$$

where  ${}_{(n)}\lambda^i$  is a unit vector tangent to  $\sigma^{,i}$ . It is known† that (9.4) is the condition that a curve of the congruence whose tangents are  ${}_{(n)}\lambda^i$  have zero first curvature. This proves one of the statements in the theorem. The converse may be demonstrated by reversing the steps of the above proof.

We now show that under the conditions of the hypothesis of Theorem 9.1, the non-null directions of the lines of curvature of the corresponding hypersurface  $\sigma = \text{const.}$  are also  $H$ -directions at  $P$ . By a change of coordinates we may write

$$(9.5) \quad \sigma = x^n, \quad g_{nn} = \frac{1}{g^{nn}} = e_n H^2(x^i), \quad g_{np} = 0.$$

From (9.5), we find

$$(9.6) \quad \sigma_{,p} = 0, \quad \sigma_{,n} = 1,$$

$$(9.7) \quad \sigma_{,np} = -\frac{H_{,p}}{H}.$$

\* In this section the indices  $p, q$  have the range  $1, 2, \dots, n-1$ .

† Eisenhart, loc. cit., p. 100.

Since the first curvature of the curve

$$(9.8) \quad x^p = \text{const.}, \quad x^n = x^n$$

is zero at  $P$ ,

$$(9.9) \quad \frac{d^2 x^i}{ds^2} + \{i|jk\} \frac{dx^j}{ds} \frac{dx^k}{ds} = 0$$

at this point. According to (9.8),

$$(9.10) \quad \frac{dx^n}{ds} = \frac{1}{H}, \quad \frac{dx^p}{ds} = 0.$$

From (9.5), (9.10), and (9.9) with  $i=q$ , we obtain  $g^{pq}H_{,p}=0$ . Since the rank of  $\|g^{pq}\|$  is  $n-1$ , this last equation is equivalent to  $H_{,p}=0$  at  $P$ . Hence, from (9.7),  $\sigma_{,np}=0$ . As a consequence of this equation and (9.6),

$$(9.11) \quad \sigma_{np} = 0$$

at the point  $P$ .

The principal directions determined by  $\sigma_{ij}$  are the vectors  $\lambda^i$  such that  $(\sigma_{ij} - \rho g_{ij})\lambda^i = 0$  for suitable  $\rho$ . It follows from (9.5) and (9.11) that the coordinate direction whose parameter is  $x^n$  is a principal direction determined by  $\sigma_{ij}$  and the vectors orthogonal to it which satisfy the equation

$$(9.12) \quad (\sigma_{pq} - \rho g_{pq})\lambda^p = 0$$

are also  $H$ -directions. Since the tensor  $\sigma_{\alpha\beta}$  defined by (8.9) for the hypersurface

$$x^\alpha = y^\alpha, \quad x^n = \text{const.}$$

is identically zero, (8.12) becomes, after a slight change in notation,

$$(9.13) \quad \sigma_{ij}\delta_\alpha^i\delta_\beta^j + e_n(e_n\Delta_1\sigma)^{1/2}\Omega_{\alpha\beta} = 0.$$

Furthermore, the tensors  $\sigma_{ij}\delta_\alpha^i\delta_\beta^j$  and  $h_{\alpha\beta}$  are, respectively, equal to  $\sigma_{pq}$  and  $g_{pq}$  after a change of notation. Since  $\Delta_1\sigma \neq 0$ , it follows from (9.12) and (9.13) that the principal directions determined by  $\sigma_{ij}$  at  $P$  other than the normal to  $\sigma = \text{const.}$  are also determined by  $\Omega_{\alpha\beta}$ . Hence they are the directions of the lines of curvature of  $\sigma = \text{const.}$  at  $P$ .

Conversely, suppose that the directions of the lines of curvature at  $P$  of  $\sigma = \text{const.}$  are all  $H$ -directions of the corresponding family of conformal geodesics of  $V_n$ . Then none of the lines of curvature at  $P$  are tangent to null vectors. In accordance with the theory of principal directions determined by the tensor  $\Omega_{\alpha\beta}$ , it follows that  $n-1$  mutually orthogonal vectors  $_{(p)}\lambda^i$  exist

at  $P$  which are tangent to lines of curvature. Hence for a proper value of  $\rho_p$

$$(9.14) \quad (\sigma_{ij} - \rho_p g_{ij})_{(p)} \lambda^i = 0.$$

If  $_{(n)}\lambda^i$  is a unit vector normal to  $\sigma = \text{const.}$  at  $P$  and

$$e_n = g_{ij} {}_{(n)}\lambda^i {}_{(n)}\lambda^j, \quad e_n \rho_n = \sigma_{ij} {}_{(n)}\lambda^i {}_{(n)}\lambda^j,$$

it follows from (9.14) that

$$(9.15) \quad (\sigma_{ij} - \rho_n g_{ij}) {}_{(n)}\lambda^i {}_{(p)}\lambda^j = 0, \quad (\sigma_{ij} - \rho_n g_{ij}) {}_{(n)}\lambda^i {}_{(n)}\lambda^j = 0.$$

Since the vectors  $_{(p)}\lambda^i, {}_{(n)}\lambda^i$  are  $n$  mutually orthogonal vectors, it follows from (9.15) that  $(\sigma_{ij} - \rho_n g_{ij}) {}_{(n)}\lambda^i = 0$ ; so the normal at  $P$  is also an  $H$ -direction of the corresponding family of conformal geodesics of  $V_n$ . This proves the following theorem:

**THEOREM 9.2.** *If the normal at a point to a hypersurface  $\sigma = \text{const.}$  is an  $H$ -direction of the corresponding family of conformal geodesics of  $V_n$ , the non-null tangents to the lines of curvature of the hypersurface at the point are also  $H$ -directions. Conversely, if at a point all the tangents to the lines of curvature of  $\sigma = \text{const.}$  are  $H$ -directions and the normal to the hypersurface at this point is a non-null vector, then the normal is also an  $H$ -direction.*

In particular, the above two theorems are true at every point of  $V_n$  only if the hypersurfaces  $\sigma = \text{const.}$  are parallel. We note that if a one-parameter family of hypersurfaces in  $V_n$  is parallel and its image in  $\bar{V}_n$  is also parallel, the family consists of the hypersurfaces  $\sigma = \text{const.}$  Since  $\Delta_1 \sigma = f(\sigma)$  if the hypersurfaces  $\sigma = \text{const.}$  are parallel, in accordance with (5.4) it is characteristic in this case that  $k_r$  has a constant value on each of these hypersurfaces.

#### IV. SOME SPECIAL QUESTIONS

**10. The Frenet equations and conformal geodesics.** The last two equations of (3.6) are equivalent to

$$(10.1) \quad \frac{d\tau_i}{ds} {}_{(3)}\xi^i = e_1 e_2 k_1 k_2, \quad \frac{d\tau_i}{ds} {}_{(1+r)}\xi^i = 0, \quad r > 2,$$

where the notation is that used in §3. We show by mathematical induction that the following equations hold for any velocity system (2.9) (in particular, for any family of conformal geodesics):

$$(10.2) \quad \begin{aligned} \frac{d^p \tau_i}{ds^p} {}_{(p+2)}\xi^i &= e_1 e_2 \cdots e_{p+1} k_1 k_2 \cdots k_{p+1}, \\ \frac{d^p \tau_i}{ds^p} {}_{(p+r)}\xi^i &= 0, \quad r > 2; p = 0, 1, \dots, n-2. \end{aligned}$$

From (3.5) and (10.1), it follows that (10.2) holds for  $p=0, 1$ . We assume that (10.2) holds for  $p=0, 1, \dots, m$ . Since  $(\delta^m \tau_i / \delta s^m)_{(m+3)} \xi^i = 0$ , we find by covariant differentiation with respect to  $s$  and use of (3.1) that

$$(10.3) \quad \frac{\delta^{m+1} \tau_i}{\delta s^{m+1}}_{(m+3)} \xi^i + \frac{\delta^m \tau_i}{\delta s^m} [-e_{m+2} k_{m+2} (m+2) \xi^i + e_{m+4} k_{m+3} (m+4) \xi^i] = 0.$$

From (10.3) and (10.2) with  $p=m$ ,

$$\frac{\delta^{m+1} \tau_i}{\delta s^{m+1}}_{(m+3)} \xi^i = e_1 e_2 \dots e_{m+2} k_1 k_2 \dots k_{m+2},$$

which is the first equation of (10.2) with  $p=m+1$ . Similarly, if we differentiate

$$\frac{\delta^m \tau_i}{\delta s^m}_{(m+1+r)} \xi^i = 0, \quad r > 2,$$

covariantly with respect to  $s$  and use (3.1), we obtain

$$\frac{\delta^{m+1} \tau_i}{\delta s^{m+1}}_{(m+1+r)} \xi^i + \frac{\delta^m \tau_i}{\delta s^m} [-e_{m+r} k_{m+r} (m+r) \xi^i + e_{m+r+2} k_{m+r+1} (m+r+2) \xi^i] = 0.$$

According to (10.2) with  $p=m$ , this equation becomes

$$\frac{\delta^{m+1} \tau_i}{\delta s^m}_{(m+1+r)} \xi^i = 0$$

which is the second equation of (10.2) with  $p=m+1$ . This completes the induction.

We apply these equations to the conformal geodesics of  $V_n$  and  $\bar{V}_n$  which are the images of the geodesics of  $\bar{V}_n$  and  $V_n$ , respectively, under the conformal transformation (2.3). Then (10.2) with  $p=0, 1$  becomes

$$(10.4) \quad \sigma_{,i} (2) \xi^i = e_1 k_1, \quad \sigma_{,i} (r) \xi^i = 0,$$

$$(10.5) \quad \frac{\delta \sigma_{,i}}{\delta s} (3) \xi^i = e_1 e_2 k_1 k_2, \quad \frac{\delta \sigma_{,i}}{\delta s} (1+r) \xi^i = 0, \quad r > 2,$$

for  $V_n$  and

$$(10.6) \quad (-\sigma)_{,i} (2) \bar{\xi}^i = \bar{e}_1 \bar{k}_1, \quad (-\sigma)_{,i} (r) \bar{\xi}^i = 0,$$

$$(10.7) \quad \frac{\delta (-\sigma)_{,i}}{\delta \bar{s}} (3) \bar{\xi}^i = \bar{e}_1 \bar{e}_2 \bar{k}_1 \bar{k}_2, \quad \frac{\delta (-\sigma)_{,i}}{\delta \bar{s}} (1+r) \bar{\xi}^i = 0, \quad r > 2,$$

for  $\bar{V}_n$ , where a notation analogous to that for  $V_n$  is used. If  $(1) \xi^i$  is chosen as corresponding to  $(1) \xi^i$  at a point  $P$ ,

$$(10.8) \quad {}_{(1)}\bar{\xi}^i = e^{-\sigma} {}_{(1)}\xi^i,$$

and it follows from (3.3), (10.4), and (10.6) that at  $P$ ,

$${}_{(2)}\bar{\xi}^i = e^{-\sigma} {}_{(2)}\xi^i, \quad \bar{k}_1 = -e^{-\sigma} k_1.$$

From (2.5) and (10.8)

$$(10.9) \quad \frac{b(-\sigma)_{,i}}{b\bar{s}} = -e^{-\sigma} \left[ \frac{b\sigma_{,i}}{b\bar{s}} - 2 \frac{d\sigma}{ds} \sigma_{,i} + \Delta_1 \sigma {}_{(1)}\xi_i \right].$$

It follows from (3.3), (10.5), (10.7), and (10.9) that at  $P$

$${}_{(3)}\bar{\xi}^i = e^{-\sigma} {}_{(3)}\xi^i, \quad \bar{k}_2 = e^{-\sigma} k_2.$$

**11. Similar families of conformal geodesics.** The families of conformal geodesics in  $V_n$  which correspond to the transformations (2.3) and

$$(11.1) \quad ds' = e^{f(\sigma)} ds, \quad f'(\sigma) \neq 0,$$

where  $f'(\sigma) = df/d\sigma$  are called *similar families of conformal geodesics*. We denote these families by  $\{C\}$  and  $\{C'\}$ , respectively. The equations analogous to (10.4) and (10.5) obtaining for  $\{C'\}$  are

$$(11.2) \quad \begin{aligned} [f(\sigma)]_{,i} {}_{(2)}\xi'^i &= e'_1 k'_1, & [f(\sigma)]_{,i} {}_{(r)}\xi'^i &= 0, \\ \frac{b[f(\sigma)]_{,i}}{b\bar{s}} {}_{(3)}\xi'^i &= e'_1 e'_2 k'_1 k'_2, & \frac{b[f(\sigma)]_{,i}}{b\bar{s}} {}_{(1+r)}\xi'^i &= 0, \quad r > 2. \end{aligned}$$

The notation in these equations is analogous to that employed in (10.4) and (10.5), the prime referring to  $\{C'\}$ . We consider curves of the two families which are tangent at a point so that  ${}_{(1)}\xi'^i = {}_{(1)}\xi^i$ . Since

$$[f(\sigma)]_{,i} = f'(\sigma) \sigma_{,i}, \quad \frac{b[f(\sigma)]_{,i}}{b\bar{s}} = f''(\sigma) \frac{d\sigma}{ds} \sigma_{,i} + f'(\sigma) \frac{b\sigma_{,i}}{b\bar{s}},$$

it follows from (3.3), (10.4), (10.5), and (11.2) that

$${}_{(2)}\xi'^i = {}_{(2)}\xi^i, \quad {}_{(3)}\xi'^i = {}_{(3)}\xi^i, \quad k'_1 = f'(\sigma) k_1, \quad k'_2 = k_2$$

at the point. Hence the ratio of the first curvatures of tangent curves of  $\{C\}$  and  $\{C'\}$  at a point is independent of their common initial direction. We also have the following theorem:

**THEOREM 11.1.** *All similar conformal geodesics which are tangent at a point of  $V_n$  have the same first and second normals and second curvatures at the point.*

In accordance with Theorem 3.1, the  $H$ -directions of  $\{C'\}$  are determined by the tensor  $f_{ij} = [f(\sigma)]_{,ij} - [f(\sigma)]_{,i} [f(\sigma)]_{,j}$ . From this equation and (4.1),

$$(11.3) \quad f_{,ij} = f'(\sigma)\sigma_{,ij} + [f''(\sigma) - f'^2(\sigma) + f'(\sigma)]\sigma_{,i}\sigma_{,j}.$$

We first assume that  $f'' - f'^2 + f' \neq 0$  and  $f' \neq 0$  at a point  $P$ . It follows from (11.3) and the remarks below (4.3) that an  $H$ -direction of  $\{C\}$  at  $P$  coincides with an  $H$ -direction of  $\{C\}'$  if and only if it is a principal direction determined by the tensor  $\sigma_{,i}\sigma_{,j}$ . But, if  $\Delta_1\sigma \neq 0$ , these principal directions are  $\sigma_{,i}$  and all vectors  $\lambda^i$  such that  $\lambda^i\sigma_{,i} = 0$ . Of course, the  $\lambda^i$  all lie in the tangent vector spaces of the hypersurfaces  $\sigma = \text{const}$ . If one of these vectors  $\lambda^i$  is an  $H$ -direction of  $\{C\}$  and therefore of  $\{C\}'$ , it follows from the second italicized statement below Theorem 8.2 that  $\lambda^i$  is tangent to a line of curvature of the hypersurface  $\sigma = \text{const}$ , passing through  $P$ . If  $\sigma_{,i}$  is a common  $H$ -direction of  $\{C\}$  and  $\{C\}'$ , it follows from Theorem 9.2 that the remaining common  $H$ -directions are the tangents of the lines of curvature of  $\sigma = \text{const}$ . at  $P$ . If the hypersurfaces  $\sigma = \text{const}$ . are parallel, this last situation is realized throughout the space.

If

$$(11.4) \quad f'' - f'^2 + f' = 0$$

and  $f' \neq 0$  at  $P$ ,  $\{C\}$  and  $\{C\}'$  have the same  $H$ -directions at this point. If (11.4) holds throughout  $V_n$ , it follows easily that

$$(11.5) \quad e^{f(\sigma)} = \frac{c_1 e^{\sigma}}{1 - c_2 e^{\sigma}},$$

where  $c_1$  and  $c_2$  are constants such that the right-hand member of (11.5) is positive. From (2.3), (11.1), and (11.5) we find that

$$(11.6) \quad \frac{a}{ds} = \frac{b}{d\bar{s}} + \frac{c}{ds'}, \quad b, c \neq 0,$$

is equivalent to (11.5).

If  $f' = 0$  at  $P$ , the  $H$ -directions of  $\{C\}'$  at this point are either all directions or  $\sigma_{,i}$  and all vectors  $\lambda^i$  such that  $\lambda^i\sigma_{,i} = 0$  according as  $f'' - f'^2 + f'$  does or does not equal zero at  $P$ . Some of these results are stated in the next theorem:

**THEOREM 11.2.** *If the hypersurfaces  $\sigma = \text{const}$ . are nonparallel, the similar families of conformal geodesics of  $V_n$  which are the images of the geodesics of  $\bar{V}_n$  and  $V'_n$  will have the same congruences of  $H$ -directions if and only if  $a/ds = b/d\bar{s} + c/ds'$ , where  $a, b (\neq 0)$ , and  $c (\neq 0)$  are constants.*

\* If  $\Delta_1\sigma = 0$  at  $P$ ,  $\sigma_{,i}$  is tangent to a null vector. In this case it is easy to show that there is a unique null vector  $\nu^i$  which is not orthogonal to  $\sigma_{,i}$  and that any vector which lies in the linear vector space determined by  $\nu^i$  and all unit vectors normal to  $\sigma_{,i}$  is a principal direction determined by  $\sigma_{,i}\sigma_{,j}$ , and conversely.

As an illustration of the above discussion as well as for its own interest, we consider the following question: What curves in  $V_n$  have principal normals equal to the principal normals of their conformal images in  $\bar{V}_n$  under the mapping (2.3)?\* According to the hypothesis and (2.4),

$$(11.7) \quad \bar{\mu}^i = e^{-\sigma} \mu^i,$$

where  $\mu^i$  and  $\bar{\mu}^i$  are the principal normals of the curve in  $V_n$  and  $\bar{V}_n$ , respectively. From (2.7) and (11.7), any curve whose principal normal is invariant under (2.3) satisfies the equation†

$$(11.8) \quad \mu^i = \frac{e^{\sigma, m}}{1 - e^{\sigma}} (g^{im} - e \xi^i \xi^m), \quad \sigma \neq 0,$$

where  $\xi^i$  is the unit tangent. Hence *the curves whose principal normals are invariant under (2.3) form a family of conformal geodesics similar to the conformal geodesics (2.8) corresponding to the given transformation.* Let (11.8) be the images of the geodesics of  $V'_n$  (determined except for a magnification). Then the induced mapping between  $V_n$  and  $V'_n$  is of the form (11.1) where

$$(11.9) \quad e^{f(\sigma)} = ae^{\sigma}/(e^{\sigma} - 1), \quad \sigma > 0, a > 0; \quad e^{f(\sigma)} = ae^{\sigma}/(1 - e^{\sigma}), \quad \sigma < 0, a > 0.$$

For the region of the coordinate space in which  $\sigma > 0$ ,‡ we investigate the curves whose principal normals in  $V_n$  and  $V'_n$  are equal under the transformation between these spaces. According to the preceding discussion, these curves are the conformal images of the geodesics of a  $V''_n$ . From (11.9), the mapping of  $V''_n$  on  $V_n$  is  $ds'' = e^{F(\sigma)} ds$ , where

$$(11.10) \quad e^{F(\sigma)} = \frac{bae^{\sigma}}{|(a-1)e^{\sigma} + 1|}, \quad b > 0.$$

Of course,  $F(\sigma) = f(f(\sigma))$  except for a magnification of  $V''_n$ . The conformal correspondence associated with (11.10) will coincide with (2.3) and  $V''_n$  with  $\bar{V}_n$  if and only if  $a=1$ ,  $b=1$ . In this case,  $V'_n$  is uniquely determined by (2.3) and from (11.9) the mapping of  $V'_n$  on  $V_n$  is given by

$$(11.11) \quad ds' = \frac{e^{\sigma}}{e^{\sigma} - 1} ds, \quad \sigma > 0.$$

Conversely, (11.11) uniquely determines  $\bar{V}_n$  and (2.3). This proves the following theorem:

\* Curves whose principal normals correspond (but not necessarily with invariant first curvatures) have been considered by V. Modesitt, loc. cit., pp. 326-328.

† At points where  $\sigma=0$ , the principal normal is invariant if and only if  $\xi^i$  is tangent to  $g^{im}\sigma_{,m}$ .

‡ In the region where  $\sigma < 0$  we simply interchange the roles of  $V_n$  and  $\bar{V}_n$ .

**THEOREM 11.3.** *Let (2.3) be a conformal transformation between a region of  $V_n$  and a region of  $\bar{V}_n$  in which  $\sigma > 0$ . Then there exists a unique Riemann space  $V'_n$  and a unique conformal mapping (11.11) of  $V'_n$  on  $V_n$  such that the images of the geodesics of  $V'_n$  have equal principal normals in  $V_n$  and  $\bar{V}_n$  and the images of the geodesics of  $\bar{V}_n$  have equal principal normals in  $V_n$  and  $V'_n$ .*

Corresponding to (11.6), we have

$$\frac{1}{ds} = \frac{1}{d\bar{s}} + \frac{1}{ds'}.$$

It follows from this equation or from the above discussion that if  $V_n$ ,  $\bar{V}_n$ , and  $V'_n$  are all subjected to the same conformal transformation,

$$ds^* = e^* ds, \quad d\bar{s}^* = e^* d\bar{s}, \quad ds'^* = e^* ds',$$

then the new spaces  $V_n^*$ ,  $\bar{V}_n^*$ , and  $V_n'^*$  may replace  $V_n$ ,  $\bar{V}_n$ , and  $V'_n$ , respectively, in Theorem 11.3. This means that the triplet of spaces  $V_n$ ,  $\bar{V}_n$ , and  $V'_n$  is a conformal triplet with respect to the property stated in Theorem 11.3.

If  $\mu'^i$  is a principal normal of a curve in  $V'_n$ , the transformation corresponding to (2.7) is

$$(11.12) \quad \mu'^i = e^{-2f(\sigma)} [\mu^i - e[f(\sigma)]_{,n} (g^{im} - e\xi^i \xi^m)],$$

where  $f(\sigma)$  is defined by (11.9) with  $a=1$ . Let  $\bar{\nu}^i$  and  $\nu'^i$  represent  $\bar{\mu}^i$  and  $\mu'^i$  considered as vectors in  $V_n$ . Then

$$(11.13) \quad \bar{\nu}^i = e^\sigma \bar{\mu}^i, \quad \nu'^i = e^{f(\sigma)} \mu'^i.$$

For any curve in  $V_n$  and its conformal images in  $\bar{V}_n$  and  $V'_n$ , we obtain from (2.7), (11.12), and (11.13) that

$$(11.14) \quad \mu^i = \bar{\nu}^i + \nu'^i.$$

In particular, if  $\nu'^i=0$ , then  $\mu^i = \bar{\nu}^i$ ; and if  $\bar{\nu}^i=0$ , then  $\mu^i = \nu'^i$ . These properties were used to define  $V'_n$ . For the geodesics of  $V_n$ ,  $\mu^i=0$ . It follows from (11.14) that the corresponding images of the geodesics of  $V_n$  in  $\bar{V}_n$  and  $V'_n$  have equal principal normals oppositely directed.

A simple computation shows that

$$e^\sigma f_{ij} + e^{f(\sigma)} \sigma_{ij} = 0,$$

so that the  $H$ -directions of the similar families of conformal geodesics (2.8) and (11.8) coincide. This also follows from the discussion preceding Theorem 10.2 since (11.9) is of the form (11.5).

**12. Conformal transformations with  $\sigma_{ij} = \phi g_{ij}$ .** In previous sections, we have seen that the conformal transformations for which the tensor  $\sigma_{ij}$  with  $\sigma$

not constant satisfies (8.21) are of a particularly simple and interesting geometric character. We now show that a very large class of  $V_n$ 's actually exists which admit such transformations. If we write  $\Omega = e^{-\sigma}$ , (8.21) becomes

$$(12.1) \quad \Omega_{,ij} = -\phi \Omega g_{ij}.$$

We first investigate the solutions of (12.1) for which

$$(12.2) \quad \Delta_1 \sigma \neq 0 \quad \text{or} \quad \Delta_1 \Omega \neq 0.$$

In this case, the equation

$$(12.3) \quad g^{ij} \Omega_{,i} \theta_{,j} = 0$$

admits  $n-1$  independent solutions\*  ${}_{(p)}\theta$ . By means of a suitable coordinate transformation, we obtain

$$(12.4) \quad {}_{(p)}\theta = x^p, \quad \Omega = x^n.$$

It follows from (12.2), (12.3), and (12.4) that

$$(12.5) \quad g^{nn} = \frac{1}{g^{nn}} \neq 0, \quad g^{pn} = 0.$$

In this coordinate system, (12.1) becomes

$$(12.6) \quad \{n | ij\} = \phi x^n g_{ij}.$$

We set  $i=n, j=p; i=p, j=q; i=n, j=n$  successively in (12.6) and use (12.5). This gives

$$\frac{\partial g_{nn}}{\partial x^p} = 0, \quad -\frac{1}{2} g^{nn} \frac{\partial g_{pq}}{\partial x^n} = \phi x^n g_{pq}, \quad \frac{1}{2} g^{nn} \frac{\partial g_{nn}}{\partial x^n} = \phi x^n g_{nn}.$$

From these equations, we find

$$(12.7) \quad \phi = \phi(x^n),$$

$$(12.8) \quad g^{nn} = - \int 2x^n \phi(x^n) dx^n,$$

$$(12.9) \quad g_{pq} = g^{nn}(x^n) h_{pq}(x^r).$$

The  $h_{pq}$  are arbitrary functions of the  $x^r$  only. Hence, a  $V_n$  admits a solution of (8.21) and (12.2) if and only if the first fundamental form of  $V_n$  may be written as

$$(12.10) \quad ds^2 = g_{pq} dx^p dx^q + g_{nn} dx^n^2,$$

\* In this section, the ranges of the indices  $p, q, r$  and  $s, t, u$  are  $1, 2, \dots, n-1$  and  $1, 2, \dots, n-2$ , respectively.

where the  $g_{ij}$  satisfy (12.8) and (12.9). Since  $\Delta_1 x^n = g^{nn}(x^n)$ , the hypersurfaces  $\sigma = \text{const.}$  are parallel. It also follows from (12.8) and (12.9)\* (or from the second italicized statement below Theorem 8.2) that these hypersurfaces have indeterminate lines of curvature.

If (12.1) and (12.2) admit other solutions  $\psi$  independent of  $x^n$ , it follows from (12.7) that  $\phi = -a$ , where  $a$  is a constant, is a necessary condition. We set  $i = n, j = p; i = n, j = n; i = p, j = q$  successively in (12.6) and use (12.5), (12.8), and (12.9). As a result, we have

$$(12.11) \quad \frac{\partial^2 \psi}{\partial x^n \partial x^p} - \frac{1}{2} \frac{\partial \psi}{\partial x^p} \frac{d \log g^{nn}}{dx^n} = 0,$$

$$(12.12) \quad \frac{\partial^2 \psi}{\partial x^{n^2}} - \frac{1}{2} \frac{\partial \psi}{\partial x^n} \frac{d \log g_{nn}}{dx^n} = a \psi g_{nn},$$

$$(12.13) \quad \frac{\partial^2 \psi}{\partial x^p \partial x^q} - \frac{\partial \psi}{\partial x^r} \{r|pq\} - \frac{\partial \psi}{\partial x^n} \{n|pq\} = a \psi g^{nn} h_{pq},$$

where in accordance with (12.8),

$$(12.14) \quad g^{nn} = ax^{n^2} + b.$$

From (12.11), we have

$$(12.15) \quad \psi = (g^{nn})^{1/2} \Lambda(x^p) + \Gamma(x^n).$$

From (12.12) and (12.15), we find

$$(12.16) \quad \frac{d^2 \Gamma}{dx^{n^2}} + \frac{d}{dx^n} (\log (g^{nn})^{1/2}) \frac{d \Gamma}{dx^n} - \frac{a}{g^{nn}} \Gamma = 0.$$

Now, by (12.5) and (12.9),

$$\{r|pq\} = g^{rs} [pq, s] = g_{nn} h^{rs} g^{nn} [pq, s]_h = \{r|pq\}_h$$

where  $[pq, s]_h$  and  $\{r|pq\}_h$  denote the Christoffel symbols of the first and second kind formed with respect to the form

$$(12.17) \quad ds^2 = h_{pq} dx^p dx^q.$$

Of course, (12.17) is the first fundamental form of each of the hypersurfaces  $\sigma = \text{const.}$  except for a magnification. Also,  $\{n|pq\} = -axg^{nn} \cdot h_{pq}$ . Substituting these results and (12.14) and (12.15) in (12.13) we have

$$(12.18) \quad \Lambda_{:pq} = \left[ ab\Lambda + a(g^{nn})^{1/2} \left( \Gamma - x^n \frac{d\Gamma}{dx^n} \right) \right] h_{pq},$$

\* Cf. Eisenhart, loc. cit., p. 182.

where the semicolon denotes covariant differentiation with respect to the form (12.17). If  $a \neq 0$ , it follows that

$$(12.19) \quad (g^{nn})^{1/2} \left( \Gamma - x^n \frac{d\Gamma}{dx^n} \right) = c,$$

where  $c$  is a constant. It is easily verified that (12.16) is a consequence of (12.19). If  $a = 0$ ,  $\Gamma = c_1 x^n + c_2$ . In both these cases,  $\Gamma$  satisfies equations similar to (12.1) where the covariant differentiation is with respect to the form (12.17). This shows that *the necessary and sufficient condition that the  $V_n$  whose first fundamental form is (12.10) admit more than one independent solution of (8.21) and (12.2) is that  $\phi$  be constant and any hypersurface  $\sigma = \text{const.}$  admit a non-constant solution of  $\Delta_{;pq} = a(b\Delta + c)h_{pq}$ .*

We now investigate the solutions of (8.21) for which

$$(12.20) \quad \Delta_1 \sigma = 0 \quad \text{or} \quad \Delta_1 \Omega = 0.$$

We first note that  $\phi = 0$  is a necessary condition for the existence of such solutions. For, according to (12.1),  $(\Delta_1 \Omega)_{,k} = g^{ij}(\Omega_{,ik}\Omega_{,j} + \Omega_{,i}\Omega_{,jk}) = -2\phi\Omega_{,k}$ . As a consequence of this equation and (12.20),  $\phi = 0$ . The equation (12.3) admits  $n-2$  independent solutions  ${}_{(n)}\theta$  besides the solution  $\Omega$ . If  ${}_{(n-1)}\theta$  is a solution of

$$(12.21) \quad g^{ij}\Omega_{,i}\theta_{,j} = 1,$$

the  $\theta$ 's and  $\Omega$  are a set of  $n$  independent variables. By means of the coordinate transformation (12.4) it follows from (12.3) and (12.21) that

$$(12.22) \quad g^{nn} = 0, \quad g^{nn} = 0, \quad g^{(n-1)n} = 1.$$

These results are equivalent to

$$(12.23) \quad g_{s(n-1)} = 0, \quad g_{(n-1)(n-1)} = 0, \quad g_{(n-1)n} = 1.$$

In this coordinate system, (12.1) becomes (12.6) with  $\phi = 0$ . It follows from (12.22) and (12.23) that (12.6) is equivalent to  $\partial g_{st}/\partial x^{n-1} = 0$ . Hence a  $V_n$  admits a nonconstant solution of (8.21) and (12.20) if and only if the first fundamental form of  $V_n$  may be written as

$$ds^2 = g_{st}(x^u, x^n)dx^s dx^t + 2dx^{n-1}dx^n + g_{nn}dx^n + g_{sn}dx^s dx^n.$$

In conclusion, we note that if  ${}_{(1)}\Omega, {}_{(2)}\Omega, \dots, {}_{(m)}\Omega$  are independent solutions of (12.1), the most general function of the  $\Omega$ 's which is also a solution of (12.1) is  $c_1 {}_{(1)}\Omega + c_2 {}_{(2)}\Omega + \dots + c_m {}_{(m)}\Omega + a$ , where the  $c$ 's are constants and  $a$  is an arbitrary constant or zero according as  $\phi$  is equal to or different from zero.

BROOKLYN COLLEGE,  
BROOKLYN, N. Y.

## BILINEAR TRANSFORMATIONS IN HILBERT SPACE\*

BY

FRANCIS J. MURRAY

**Introduction.** A function of two variables  $h = F(f, g)$ , where  $h, f$ , and  $g$  are all elements of Hilbert Space may be termed a bilinear transformation if it is linear in  $f$  and linear in  $g$ . A more formal definition is given in §1. While a complete treatment of bilinear transformations would obviously require a very lengthy discussion, we wish to point out in this paper that many of the methods used in the study of linear transformations are applicable to them, with, of course, certain modifications. Many elementary notions can be extended and corresponding results obtained. For certain classes of bilinear transformations, there is even a "canonical resolution" (cf. §5, Theorem 7).

Bilinear transformations have appeared in the work of Kerner.† While the first Fréchet differential is a linear transformation, the second is bilinear, and it is this connection which was studied by Kerner. We shall show the relationship between bilinear transformations and rings of operators.‡

Mazur and Orlicz have pointed out the relationship between bilinear (and multilinear) transformations and polynomial transformations (cf. [5], p. 59). Polynomial transformations have also been studied by Banach (cf. [2]). We shall have occasion to use some of their results.

There is a very simple relationship between bilinear transformations and trilinear forms. For instance, if  $F(f, g)$  is a bilinear transformation, then

$$\alpha(f, g, h) = (F(f, g), h),$$

(,) denoting the inner product, is linear in  $f$  and  $g$ , conjugate linear in  $h$ . For finite dimensional spaces, trilinear and multilinear forms have been discussed by Hitchcock and also by Oldenburger (cf. [3], [12], [13]). While a study of the infinite case demands more abstract methods and a decided shift in emphasis, nevertheless there is a certain similarity in the ideas involved in Hitchcock's paper and our discussion.

The results of §§1-5 can readily be extended to multilinear forms. In connection with this, it should be pointed out that in general we have a certain freedom in considering the nature of  $T$  (§4), in regard to the spaces on which it operates. For instance, if  $F$  is trilinear, we may consider  $T_1, T_2, T_3$  defined by

\* Presented to the Society, February 26, 1938; received by the editors June 29, 1938.

† Cf. [4]. Numerals in brackets refer to the bibliography at the end of this paper.

‡ Rings of operators are discussed in [9], [7], and [8].

$$\begin{aligned}(F(f, g, h), k) &= (T_1 f \otimes g \otimes h, k) = (T_2 f \otimes g, h \otimes k) \\ &= (T_3 f, g \otimes h \otimes k).\end{aligned}$$

Thus, if in the general case  $T_1$  is a transformation from a  $k_1$ -multiple to a  $k_2$ -multiple space and  $T_2$  is a transformation from a  $k_2$ -multiple to a  $k_3$ -multiple space, then  $T_2 T_1$  is a transformation from a  $k_1$ -multiple to a  $k_3$ -multiple space and so can be regarded as determining a  $(k_1 + k_3 - 1)$ -linear transformation. This last process corresponds to the notion of "composition" discussed by Oldenburger in [12].

In §1 we define a bilinear transformation and consider various possibilities for its domain. In §2 the notion of continuity and of a matrix for a bilinear transformation is discussed. Closure is discussed in §3 and its relationship with continuity. In §4 the "hypergraph" is discussed, and we show the connection between completely linear transformations  $F(f, g)$  and linear transformations between  $\mathfrak{H} \otimes \mathfrak{H}$  and  $\mathfrak{H}$ . In §5 the hyper-properties of completely linear transformations are discussed.

In §§6-12 we discuss the possibility of regarding Hilbert space as a hypercomplex number system. In §6 it is shown that this requires the introduction of a bilinear transformation  $F(f, g)$  which is associative; that is, one for which  $F(f, F(g, h)) = F(F(f, g), h)$ . In §7 it is shown that if such an  $F(f, g)$  is closed with respect to adjoints (Definition 7.2), then there is associated with it an algebraic ring of operators  $\mathbf{M}$ . If  $F$  also has certain continuity properties, then there is an element  $f_0$ , such that the operation  $E_0$ , defined by the equation  $E_0 f = F(f_0, f)$ , is a projection and is also the maximal idempotent for  $\mathbf{M}$ . The operation  $E_0$  and the possible  $f_0$ 's are discussed further in §8. In §9 we discuss conditions which are sufficient for an  $\mathbf{M}$  and an  $f_0$  to determine an associative bilinear transformation  $F$ .

In §10 examples of the foregoing are cited, and we also discuss the case in which an associative bilinear transformation is everywhere defined. In §11 for  $\mathbf{M}$  a ring of operators, we consider the relationship between  $F$  and  $\mathbf{M}'$ . In §12 we deal with the abelian cases of associative bilinear transformations.

Certain further examples are appended.

The second part of this paper represents also a development in a certain direction of a recent joint paper of J. von Neumann and the writer [8]. While the proofs given here were obtained independently, it is impossible to evaluate to what extent the general outline of the theory and the related notions have been influenced by previous discussions with Professor von Neumann. The author is also indebted to the referee for many suggestions in both parts of the present paper.

1. We introduce first the following definition:

DEFINITION 1.1. A function  $F(f, g)$  of pairs of elements of Hilbert space is said to be a bilinear transformation if the following conditions hold:

- (a) The values of  $F(f, g)$  are in Hilbert space.
- (b) If  $g$  is such that there is an  $f$  for which  $F(f, g)$  is defined, then  $R_g f = F(f, g)$  is a linear transformation on  $f$ .
- (c) If  $f$  is such that there is a  $g$  for which  $F(f, g)$  is defined, then  $T_f g = F(f, g)$  is a linear transformation on  $g$ .

The "graph" has yielded effective methods for the study of linear transformations (cf. [11]). For linear transformations, the method of procedure is to form\*  $\mathfrak{H} \oplus \mathfrak{H}$  and then to consider the set of pairs  $\{f, Tf\}$  in this space. This set is called the graph. The statement that  $T$  is linear is equivalent to the statement that the graph is a linear manifold. The usefulness of the graph depends upon this fact.

Inasmuch as pairs of elements are involved, one might attempt to obtain a graph for bilinear transformations by forming  $(\mathfrak{H} \oplus \mathfrak{H}) \oplus \mathfrak{H}$  and considering the elements of this space which are in the form  $\{\{f, g\}, F(f, g)\}$ . However the essential linearity property of  $F(f, g)$  is the property that

$$F(f_1, g) + F(f_2, g) = F(f_1 + f_2, g), \quad F(f, g_1) + F(f, g_2) = F(f, g_1 + g_2).$$

We would therefore demand of the graph that

$$\{\{f_1, g\}, F(f_1, g)\} + \{\{f_2, g\}, F(f_2, g)\} = \{\{f_1 + f_2, g\}, F(f_1 + f_2, g)\}.$$

However the left-hand sum is

$$\{\{f_1, g\} + \{f_2, g\}, F(f_1, g) + F(f_2, g)\} = \{\{f_1 + f_2, 2g\}, F(f_1 + f_2, g)\}$$

by the usual rules for addition in the space  $(\mathfrak{H} \oplus \mathfrak{H}) \oplus \mathfrak{H}$ . In general, therefore, the desired equation does not hold.

This difficulty is easily traced to the fact that the linearity properties of  $\{f, g\}$  are not the same as those of  $F(f, g)$ . However in the space  $\mathfrak{H} \otimes \mathfrak{H}$  (cf. [7], loc. cit.) the expression  $f \otimes g$  has precisely the same linearity properties as  $F(f, g)$ . As a consequence the elements of  $(\mathfrak{H} \otimes \mathfrak{H}) \oplus \mathfrak{H}$  in the form  $\{f \otimes g, F(f, g)\}$  are readily seen to represent the linearity properties of  $F(f, g)$ . It seems expedient therefore, to propose the following definition:

DEFINITION 1.2. The graph  $\mathfrak{G}$  of a bilinear transformation is that set of elements of  $(\mathfrak{H} \otimes \mathfrak{H}) \oplus \mathfrak{H}$  in the form  $\{f \otimes g, h\}$  for which  $h = F(f, g)$ . The domain of  $F(f, g)$  is the set of elements  $f \otimes g$  of  $\mathfrak{H} \otimes \mathfrak{H}$  for which  $F(f, g)$  is defined.

In Definition 1.1, two pairs  $\{f, g\}, \{f^{(1)}, g^{(1)}\}$  are considered distinct un-

\* The operations  $\oplus$  and  $\otimes$  as applied to Hilbert space, have been discussed in [14], Theorem 1.26, and [7], chap. 2, respectively.

less  $f=f^{(1)}$ ,  $g=g^{(1)}$ . But  $f \otimes g = f^{(1)} \otimes g^{(1)}$  does not imply  $f=f^{(1)}$ ,  $g=g^{(1)}$ . However the situation is clarified by the following discussion. We begin with the statement that if  $f \otimes g = f^{(1)} \otimes g^{(1)}$  and if  $F$  is a bilinear transformation for which  $F(f, g)$  and  $F(f^{(1)}, g^{(1)})$  are defined, then  $F(f, g) = F(f^{(1)}, g^{(1)})$ .

Two cases arise. If  $f \otimes g = 0$ , then  $\|f \otimes g\| = \|f\| \cdot \|g\| = 0$ , and either  $f=0$  or  $g=0$ . In the first case  $F(0, g) = R_g 0 = 0$ . Similarly  $F(f, 0) = 0$ . Thus  $f \otimes g = 0$  implies  $F(f, g) = 0$ . Therefore  $f \otimes g = f^{(1)} \otimes g^{(1)} = 0$  implies that  $F(f, g) = 0 = F(f^{(1)}, g^{(1)})$ .

The case  $f \otimes g \neq 0$  is shown by first proving that if  $f \otimes g = f^{(1)} \otimes g^{(1)} \neq 0$ , then  $f^{(1)} = \lambda f$ ,  $g^{(1)} = \mu g$  and  $\lambda\mu = 1$ . Under these circumstances both  $f$  and  $g^{(1)}$  are not zero. It follows then that if we orthonormalize  $f$ ,  $f^{(1)}$  by the Gram-Schmidt process, we obtain either one,  $\phi$ , or two,  $\phi_1$  and  $\phi_2$ , orthonormal elements. If the latter case could arise, then  $f = a\phi_1$ ,  $f^{(1)} = b_1\phi_1 + b_2\phi_2$  with  $b_2 \neq 0$ . Then  $f \otimes g = f^{(1)} \otimes g^{(1)}$  may be written as

$$\phi_1 \otimes ag = \phi_1 \otimes b_1g^{(1)} + \phi_2 \otimes b_2g^{(1)}.$$

The argument in [7], §2.4, now implies that  $b_2g^{(1)} = 0$ . But since both  $b_2$  and  $g^{(1)}$  are not zero, this is impossible; hence only one orthonormal element  $\phi$  can arise.

Thus  $f = a\phi$ ,  $f^{(1)} = b\phi$ , and  $ab \neq 0$ . Hence  $f^{(1)} = (b/a)f$ . Also [7], §2.4, can now be used to show that  $ag = bg^{(1)}$  or  $g^{(1)} = (a/b)g$ . Thus  $f \otimes g = f^{(1)} \otimes g^{(1)} \neq 0$  implies  $f^{(1)} = \lambda f$ ,  $g^{(1)} = \mu g$ , and  $\lambda\mu = 1$ . Consequently

$$F(f^{(1)}, g^{(1)}) = F(\lambda f, \mu g) = \lambda F(f, \mu g) = \lambda\mu F(f, g) = F(f, g).$$

These results show that while a pair  $\{f \otimes g, h\}$  in the graph may represent more than one equation  $h = F(f, g)$ ; nevertheless (except for  $f \otimes g = 0$ ) each represented equation is a consequence of any other due to the nature of  $F$ .

Notice that it follows from Definition 1.1 that the set  $\mathfrak{N}_L$  of  $f$ 's for which  $F(f, 0)$  is defined must be a linear manifold since  $F(f, 0) = R_0 f$ . Furthermore, it will contain the set  $\mathfrak{A}_L$  of all  $f$ 's for which  $F(f, g)$  is defined for a nonzero  $g$ . We assume that  $\mathfrak{N}_L$  is precisely the linear manifold determined by  $\mathfrak{A}_L$ , unless an explicit extension is made. This, then, is the sense in which  $0 = f \otimes 0$  is to be understood as in the domain of  $F$ .

We next discuss various possibilities for the domain of  $F(f, g)$ .

DEFINITION 1.3. The domain of  $F(f, g)$  is said to be dense if it is dense in the set of  $f \otimes g$  of the space  $\mathfrak{S} \otimes \mathfrak{S}$ .

DEFINITION 1.4. The domain of  $F(f, g)$  is said to be rectangular if, whenever it contains  $f_1 \otimes g_1$  and  $f_2 \otimes g_2$  both different from zero, it also contains  $f_1 \otimes g_2$  and  $f_2 \otimes g_1$ .

**DEFINITION 1.5.** The domain of a bilinear transformation  $F(f, g)$  is said to be completely linear if with  $f_i \otimes g_i$ , ( $i=1, 2, \dots, n$ ), it also contains every element  $f \otimes g$  such that  $f \otimes g = \sum_{i=1}^n a_i f_i \otimes g_i$ .

**LEMMA 1.1.** If the domain of a bilinear transformation  $F(f, g)$  is rectangular, then the domain is completely linear.

**Proof.** Suppose that  $f_1 \otimes g_1, f_2 \otimes g_2, \dots, f_n \otimes g_n$  are in the domain of  $F(f, g)$  and  $f \otimes g = \sum_{i=1}^n a_i f_i \otimes g_i \neq 0$ . Putting  $a_0 = -1$ ,  $f_0 = f$ ,  $g_0 = g$ , we may write  $\sum_{i=0}^n a_i f_i \otimes g_i = 0$ . Let us orthonormalize  $g_0, g_1, \dots, g_n$  by the Gram-Schmidt process, and let the result be  $\phi_0, \phi_1, \dots, \phi_k$ . Since  $f_0 \otimes g_0 \neq 0$ ,  $g_0 \neq 0$ . Hence  $C\phi_0 = g_0$ ,  $C \neq 0$ . Also for  $i=1, 2, \dots, n$ ,  $g_i = \sum_{j=0}^k b_{i,j} \phi_j$ . Substituting we get

$$\begin{aligned} 0 &= \sum_{i=0}^n a_i f_i \otimes g_i = \sum_{i=0}^n \sum_{j=0}^k a_i b_{i,j} f_i \otimes \phi_j \\ &= \sum_{j=0}^k \left( \sum_{i=0}^n a_i b_{i,j} f_i \right) \otimes \phi_j \\ &= \left( -Cf + \sum_{i=1}^n a_i b_{i,0} f_i \right) \otimes \phi_0 + \sum_{j=1}^k \left( \sum_{i=0}^n a_i b_{i,j} f_i \right) \otimes \phi_j. \end{aligned}$$

This implies by [7], §2.4, that  $-Cf + \sum_{i=1}^n a_i b_{i,0} f_i = 0$  or  $f = \sum_{i=1}^n c_i f_i$ . Now since the domain of  $F(f, g)$  is rectangular, it must contain  $f_i \otimes g_1$ ; hence by Definition 1.1 it must also contain  $\sum_{i=1}^n c_i f_i \otimes g_1 = f \otimes g_1$ .

A similar proof will show that  $f_1 \otimes g$  is also in the domain of  $F(f, g)$ , and these two results imply, by Definition 1.2, that  $f \otimes g$  is in the domain.

The converse of this lemma is not true (cf. Example 1 below).

**DEFINITION 1.6.** The domain of  $F(f, g)$  is said to be rectangularly dense if, for a dense set of  $g$ 's in  $\mathfrak{S}$ ,  $R_g$  has a dense domain and if for a dense set of  $f$ 's in  $\mathfrak{S}$ ,  $T_f$  has a dense domain (cf. Definition 1.1 above).

It is easily seen that if the domain of  $F(f, g)$  is rectangular and dense, it is rectangularly dense. The converse does not hold. The less restrictive condition is sufficient for some purposes.

**DEFINITION 1.7.** The domain of  $F(f, g)$  is said to be symmetric, if with  $f \otimes g$  it contains  $g \otimes f$ .

2. We next discuss certain elementary properties which  $F$  may have.

**DEFINITION 2.1.** A bilinear transformation  $F(f, g)$  is said to be continuous at a point  $f_0 \otimes g_0$  of its domain if, given any  $\epsilon > 0$ , we can find a  $\delta > 0$ , such that when  $f \otimes g$  is in the domain of  $F$  and  $\|f - f_0\| < \delta$ ,  $\|g - g_0\| < \delta$ , then  $\|F(f, g) - F(f_0, g_0)\| < \epsilon$ . The transformation  $F(f, g)$  will be said to be continuous if it is continuous at every point of its domain.

This definition is, of course, the usual one for continuity in two variables. The notion of neighborhood implicit in it is equivalent to the neighborhood notion in  $\mathfrak{H} \oplus \mathfrak{H}$ . On the other hand, the additive properties of  $F(f, g)$  are, as we shall see, best described in terms of  $\mathfrak{H} \otimes \mathfrak{H}$ . Despite this disharmony many results concerning the relation of additivity and continuity can be proved.

**THEOREM 1.** *If  $F(f, g)$  has a rectangular domain and is continuous at one point, then  $F(f, g)$  is continuous and there exists a constant  $C$  such that  $\|F(f, g)\| \leq C\|f\| \cdot \|g\|$ .*

It is a consequence of [5] (§11, p. 179) and Principle A (p. 59) that continuity at a single point implies continuity at every point. Thus  $F(f, g)$  is continuous.

In particular,  $F(f, g)$  is continuous at the origin. A familiar process used in [14] for the proof of Theorem 2.21, yields the existence of a  $C$  such that  $\|F(f, g)\| \leq C\|f\| \cdot \|g\|$ .

As with linear transformations, there exists a matrix theory for bilinear forms. The following theorem expresses this fact:

**THEOREM 2.** *Let  $F(f, g)$  be bilinear. Then if  $\phi_1, \phi_2, \dots$  is a complete orthonormal set in  $\mathfrak{H}$ , there exists a set of bilinear complex-valued functions  $\alpha_i(f, g)$  such that*

$$F(f, g) = \sum_{i=1}^{\infty} \alpha_i(f, g) \phi_i$$

*for every  $f \otimes g$  in the domain of  $F$ . If  $F$  is bounded, there exists for each  $i$  a bounded conjugate linear transformation  $T_i$  such that  $\alpha_i(f, g) = (f, T_i g)$ .*

**Proof.** The first statement is shown by taking  $\alpha_i(f, g) = (F(f, g), \phi_i)$  and applying Theorem 1.9 of [14]. If  $F$  has the bound  $C$ , then,

$$|\alpha_i(f, g)| = |(F(f, g), \phi_i)| \leq \|F(f, g)\| \cdot \|\phi_i\| \leq C\|f\| \cdot \|g\|.$$

A proof, very similar to that of Theorem 2.28 of [14] will now show the existence of  $T_i$  and its boundness under these circumstances.

3. We make the following definition:

**DEFINITION 3.1.** *If the graph  $\mathfrak{F}$  of  $F(f, g)$  is closed, then  $F(f, g)$  is said to be closed.*

Closure of the graph  $\mathfrak{F}$  is equivalent to the following statement concerning  $F$ : If  $f_n \otimes g_n \rightarrow f \otimes g$  and  $F(f_n, g_n) \rightarrow h$ , then  $f \otimes g$  is in the domain of  $F$  and  $F(f, g) = h$ .

We have the following relationship between closure and continuity for bilinear transformations:

**THEOREM 3.** *If  $F(f, g)$  is defined for every pair  $f \otimes g$  and is closed, then  $F(f, g)$  is bounded.*

**Proof.** If we keep  $g$  fixed,  $F(g, f) = T_g f$  is, under these hypotheses, a closed linear transformation with domain  $\mathfrak{H}$ . Hence by a well known result (cf. [1], chap. 3, Theorem 7, p. 41)  $T_g$  is bounded, with a bound  $C_g$ .

Similarly, if we keep  $f$  fixed,  $F(g, f) = R_f g$  defines a bounded linear transformation  $R_f$ .

Now, there must be a neighborhood of 0 in  $\mathfrak{H}$  for which  $C_g$  is bounded. For otherwise, we can find a sequence  $g_n$  such that  $g_n \rightarrow 0$  and  $C_{g_n} \rightarrow \infty$  as  $n \rightarrow \infty$ . However for each  $f$ ,

$$\lim_{n \rightarrow \infty} T_{g_n} f = \lim_{n \rightarrow \infty} F(g_n, f) = \lim_{n \rightarrow \infty} R_f g_n = R_f 0 = 0.$$

Thus the  $T_{g_n}$  are a convergent sequence of bounded transformations; so by [1] (chap. 5, Theorem 5, p. 80) they are uniformly bounded. This contradicts the assumption that  $C_{g_n} \rightarrow \infty$  as  $n \rightarrow \infty$ .

Thus there exist positive constants  $k$  and  $\delta$ , such that  $\|g\| \leq \delta$  implies  $C_g \leq k$ . Since  $g = (\|g\|/\delta)g'$  for some  $g'$  with  $\|g'\| = \delta$ , we have

$$\|F(g, f)\| = (\|g\|/\delta)\|F(g', f)\| = (\|g\|/\delta)\|T_{g'} f\| \leq (k/\delta) \cdot \|g\| \cdot \|f\|.$$

This completes the proof of Theorem 3.

4. In the preceding sections, we dealt with the properties of  $F(f, g)$  which are concerned with the graph  $\mathfrak{F}$ . This graph determines a linear manifold  $l(\mathfrak{F})$  in  $(\mathfrak{H} \otimes \mathfrak{H}) \oplus \mathfrak{H}$ . We now consider  $l(\mathfrak{F})$ . First we make a definition as follows:

**DEFINITION 4.1.** *A bilinear transformation  $F(f, g)$  will be said to be completely linear provided that the domain of  $F(f, g)$  is completely linear and the relationship  $f \otimes g = \sum_{i=1}^n a_i f_i \otimes g_i$  among the elements of the domain implies*

$$F(f, g) = \sum_{i=1}^n a_i F(f_i, g_i).$$

In the relationship  $f \otimes g = \sum_{i=1}^n a_i f_i \otimes g_i$  we need only consider the case in which  $f \otimes g \neq 0$ . For if  $f \otimes g = 0$  and at least one of the  $a_i f_i \otimes g_i \neq 0$ , then by applying the above definition to  $n-1$  elements  $f_i \otimes g_i$ , we see that the equation on the values of  $F$  is still fulfilled. If all the  $a_i f_i \otimes g_i = 0$ , the same result obtains. Hence the above definition is equivalent to the corresponding one in which the condition  $f \otimes g \neq 0$  is added.

It is important to note that a bilinear transformation is not necessarily completely linear, as we show by Example 2 below. The value of the notion of complete linearity lies in the following theorem:

**THEOREM 4.** *If a bilinear transformation  $F(f, g)$  is completely linear,  $l(\mathfrak{F})$  is the graph of a linear transformation  $T$  from  $\mathfrak{S} \otimes \mathfrak{S}$  to  $\mathfrak{S}$  (cf. [6], Definition 1.2, p. 303). Also  $T$  is such that  $f \otimes g$  is in the domain of  $T$  if and only if it is in the domain of  $F(f, g)$ ; and when this occurs,  $F(f, g) = Tf \otimes g$ .*

**Proof.** Since  $l(\mathfrak{F})$  is linear, it will be the graph of a transformation from  $\mathfrak{S} \otimes \mathfrak{S}$  to  $\mathfrak{S}$  if and only if  $\{0, h\} \in l(\mathfrak{F})$  implies  $h=0$ . Now if  $\{k, h\}$  is in  $l(\mathfrak{F})$ , then there must exist a finite number of elements  $f_i \otimes g_i$  in the domain of  $F$  and constants  $a_1, a_2, \dots, a_n$ , such that

$$k = \sum_{i=1}^n a_i f_i \otimes g_i, \quad h = \sum_{i=1}^n a_i F(f_i, g_i).$$

If  $k=0$ , then, as we have remarked after Definition 4.1 above, complete linearity implies  $h=0$ . This proves the first statement of the theorem.

The remaining statements are immediate consequences of the definition of  $l(\mathfrak{F})$ , Definitions 1.5 and 4.1 above, and Definition 1.2 of [6].

The converse of this is the following, the proof of which we omit:

**THEOREM 5.** *If  $T$  is a linear transformation from  $\mathfrak{S} \otimes \mathfrak{S}$  to  $\mathfrak{S}$ , the equation  $Tf \otimes g = F(f, g)$  determines a completely linear bilinear transformation  $F(f, g)$ , with domain the set of  $f \otimes g$  which are in the domain of  $T$ .*

It is an immediate consequence of Definition 1.3 and the fact that the set of  $f \otimes g$ 's spans  $\mathfrak{S} \otimes \mathfrak{S}$  that if the domain of  $F$  is dense and  $T$  exists, then the domain of  $T$  is dense.

**THEOREM 6.** *If the domain of a bilinear transformation  $F(f, g)$  is rectangular,  $F$  is completely linear.*

**Proof.** We know from Lemma 1.1, that the domain of  $F(f, g)$  is completely linear. Suppose now that  $f_1 \otimes g_1, f_2 \otimes g_2, \dots, f_n \otimes g_n$  are in the domain of  $F(f, g)$  and  $f \otimes g = \sum_{i=1}^n a_i f_i \otimes g_i$ . We must show that

$$F(f, g) = \sum_{i=1}^n a_i F(f_i, g_i).$$

Now by letting  $a_0 = -1, f_0 = f, g_0 = g$ , we see that this is a consequence of the statement that the relationship  $\sum_{i=0}^n a_i f_i \otimes g_i = 0$ , among elements in the domain of  $F$ , implies  $\sum_{i=0}^n a_i F(f_i, g_i) = 0$ .

We shall show this last statement inductively with respect to  $n$ . If  $n=0$  and  $a_0 f_0 \otimes g_0 = 0$ , then  $a_0$  or  $f_0$  or  $g_0$  is zero and hence  $a_0 F(f_0, g_0) = 0$  by Definition 1.1. Now suppose it is true for  $n-1$ ; we shall show it for  $n$ . We can suppose that  $g_0 \neq 0$ , since otherwise we have a situation equivalent to that of the case for  $n-1$ .

We can now proceed as in the proof of Lemma 1.1 to orthonormalize the  $g_0, g_1, \dots, g_n$ , obtaining  $\phi_0, \phi_1, \dots, \phi_k$  with  $g_i = \sum_{j=0}^k b_{i,j} \phi_j$ , ( $i=0, 1, \dots, n$ ). We then obtain, as there, that

$$0 = \sum_{i=0}^n a_i f_i \otimes g_i = \sum_{j=0}^k \left( \sum_{i=0}^n a_i b_{i,j} f_i \right) \otimes \phi_j;$$

and [7], §2.4, implies that

$$\sum_{i=0}^n a_i b_{i,j} f_i = 0, \quad j = 0, 1, \dots, k.$$

By Definition 1.4,  $f_i \otimes g_j$  is in the domain of  $F$  for  $i, j=0, 1, \dots, n$ . Since the  $\phi_j$ 's are linear combinations of the  $g_j$ 's, Definition 1.1 yields that  $f_i \otimes \phi_j$  is in the domain of  $F$  for  $i=0, 1, \dots, n$  and  $j=0, 1, \dots, k$ . Thus

$$\begin{aligned} \sum_{i=0}^n a_i F(f_i, g_i) &= \sum_{i=0}^n \sum_{j=0}^k a_i b_{i,j} F(f_i, \phi_j) \\ &= \sum_{j=0}^k F \left( \sum_{i=0}^n a_i b_{i,j} f_i, \phi_j \right) = \sum_{j=0}^k F(0, \phi_j) = 0. \end{aligned}$$

5. We deal in this section with completely linear  $F(f, g)$ .

DEFINITION 5.1. A completely linear  $F(f, g)$  is said to be hypercontinuous if the  $T$  of Theorem 4 is continuous.

Since  $T$  is linear, the known results on linear transformations are immediately applicable here. For instance, we might define hypercontinuity at a point for  $F(f, g)$  as continuity for  $T$  at the point  $f \otimes g$  of  $\mathfrak{S} \otimes \mathfrak{S}$ . Then hypercontinuity at a point implies hypercontinuity. The correspondence of hypercontinuity and hyperboundness also results.

DEFINITION 5.2. A completely linear  $F(f, g)$  is said to be hyperclosable if  $T$  possesses a closed extension  $[T]$ . Let  $\bar{F}(f, g) = [T]f \otimes g$ , as in Theorem 5.

The transformations  $F$  and  $\bar{F}$  are not, in general, equal, even if  $F$  is closed and possesses a rectangular and symmetric domain (cf. Example 3 below).

It should be pointed out that while it is obvious that the hyper-properties imply the corresponding simple properties of  $F$ , the converse is not true. Example 4 below is an example of an  $F(f, g)$  which is bounded, has as domain all pairs  $f \otimes g$ , and yet is not even hyperclosable.

We next discuss hyperclosable transformations.

THEOREM 7. Let  $F$  be hyperclosable and  $F = \bar{F}$ . There exists a self-adjoint transformation  $H$  on  $\mathfrak{S} \otimes \mathfrak{S}$  and a partially isometric transformation  $W$  from  $\mathfrak{S} \otimes \mathfrak{S}$  to  $\mathfrak{S}$  such that the domain of  $F$  is exactly the set of  $f \otimes g$  in the domain of  $H$

and  $F(f, g) = WHf \otimes g$ . To each such  $H$  and  $W$  with the same zero manifold we can find an  $F$  with  $F = \bar{F}$  such that  $F(f, g) = WHf \otimes g$ .

The canonical resolution of  $[T]$  (cf. [6], Theorem 1.24, p. 312) yields this result immediately.

**THEOREM 8.** *If  $F(f, g)$  is hyperclosable and  $F = \bar{F}$ , then the orthonormal set  $\phi_1, \phi_2, \dots$  of Theorem 2 can be chosen so that for each  $i$  there exists a conjugate linear transformation  $T_i$  of finite norm such that*

$$\alpha_i(f, g) = (f, T_i g).$$

**Proof.** Since  $[T]$  is closed, we can by [6] (Theorem 6, p. 315) determine a set of mutually orthogonal closed linear manifolds  $\mathfrak{D}_i$ , ( $i=0, \pm 1, \pm 2, \dots$ ), in  $\mathfrak{S} \otimes \mathfrak{S}$  and a similar set  $\mathfrak{R}_i$  in  $\mathfrak{S}$  such that  $[T]$  takes  $\mathfrak{D}_i$  into  $\mathfrak{R}_i$ . Let us choose in each  $\mathfrak{R}_i$  an orthonormal set  $\phi_{i,1}, \phi_{i,2}, \dots$ , complete in  $\mathfrak{R}_i$ .

From the definition of  $\mathfrak{D}_i$  and  $\mathfrak{R}_i$  in [6], Theorem VI, by means of [6], Theorem IV (we take  $F' = WFW^*$ ) and [6], Theorem 1.24, we see that  $\mathfrak{R}_i$  is in the domain of  $[T]^*$ . Hence, for all  $\bar{f}$  in the domain of  $[T]$ ,

$$[T]\bar{f} = \sum_{i=-\infty}^{\infty} \sum_{j=1}^{\infty} ([T]\bar{f}, \phi_{i,j}) \phi_{i,j} = \sum_{i=-\infty}^{\infty} \sum_{j=1}^{\infty} (\bar{f}, [T]^* \phi_{i,j}) \phi_{i,j}.$$

Now rearranging the  $\phi_{i,j}$ 's into a single sequence  $\{\psi_k\}$  and letting  $\bar{g}_k = [T]^* \phi_{i,j} = [T]^* \psi_k$ , we obtain that for all  $\bar{f}$  in the domain of  $[T]$

$$[T]\bar{f} = \sum_{i=1}^{\infty} (f, \bar{g}_i) \psi_i.$$

We next choose a complete orthonormal sequence  $\{\phi_l\}$  in  $\mathfrak{S}$ . Then

$$\bar{g}_i = \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} a_{l,k}^{(i)} \phi_l \otimes \phi_k$$

with

$$\sum_{l=1}^{\infty} \sum_{k=1}^{\infty} |a_{l,k}^{(i)}|^2$$

finite. Let  $\bar{f} = f \otimes g$  be in the domain of  $[T]$ . Since  $F = \bar{F}$ , it is also in the domain of  $T$ . Let  $f = \sum_{i=1}^{\infty} x_i \phi_i$ ,  $g = \sum_{j=1}^{\infty} y_j \phi_j$ . Let  $T_i$  be the conjugate linear transformation such that

$$T_i g = \sum_{l=1}^{\infty} \left( \sum_{j=1}^{\infty} a_{l,j}^{(i)} \bar{y}_j \right) \phi_l.$$

Then  $T_i$  is of finite norm and

$$\begin{aligned}
 (f, T_i g) &= \sum_{l=1}^{\infty} x_l \left( \sum_{j=1}^{\infty} \bar{a}_{l,i}^{(i)} y_j \right) = \sum_{l=1}^{\infty} \sum_{j=1}^{\infty} x_l y_j \bar{a}_{l,i}^{(i)} \\
 &= \left( \sum_{l=1}^{\infty} \sum_{j=1}^{\infty} x_l y_j \phi_l \otimes \phi_j, \sum_{l=1}^{\infty} \sum_{j=1}^{\infty} a_{l,i}^{(i)} \phi_l \otimes \phi_j \right) \\
 &= (f \otimes g, \bar{g}_i) = \alpha_i(f, g).
 \end{aligned}$$

This proves the theorem.

Theorems 2 and 8 indicate that the relation between the set of hyperclosable and that of the bounded bilinear transformations is analogous to that between the linear transformations of finite norms and the bounded transformations.

6. An arbitrary vector space becomes a hypercomplex number system if a "rule of multiplication"  $f \times g$  is given. Addition in the system is usually the vector sum  $f + g$  of the original space. If, in particular, we consider Hilbert space, we are led to consider the ways in which a "multiplication rule"  $\times$  may be defined on the space.

Such an  $\times$  operation should be distributive; that is, we should have  $(f_1 + f_2) \times g = f_1 \times g + f_2 \times g$  and  $f \times (g_1 + g_2) = f \times g_1 + f \times g_2$ . It should also be such that  $af \times bg = ab(f \times g)$ . So if we hold  $f$  fixed, we see that  $T_f g = f \times g$  is a linear transformation on  $g$ . Similarly when  $g$  is fixed,  $R_g f = f \times g$  is a linear transformation of  $f$ . Thus  $F(f, g) = f \times g$  is a bilinear transformation in Hilbert space.

The associativity property,  $f \times (g \times h) = (f \times g) \times h$ , is also desired. For the corresponding bilinear transformation this would mean that  $F(f, F(g, h)) = F(F(f, g), h)$ . Since, however, it is too restrictive to demand that  $F(f, g)$  be defined for every  $f$  and  $g$ , we define associativity for bilinear transformations as follows:

**DEFINITION 6.1.** A bilinear transformation  $F(f, g)$  is said to be associative if it satisfies the following conditions:

- For every  $g$ ,  $F(f, g)$  is defined for every  $f$  in a linear set  $\mathfrak{A}$ .
- $T_f g = F(f, g)$ , ( $f \in \mathfrak{A}$ ), is a bounded linear transformation.
- If  $f_1$  and  $f_2$  are in  $\mathfrak{A}$ ,  $F(f_1, f_2)$  is in  $\mathfrak{A}$ .
- For  $f_1$  and  $f_2 \in \mathfrak{A}$  and  $g \in \mathfrak{S}$ ,

$$F(f_1, F(f_2, g)) = F(F(f_1, f_2), g).$$

Thus if  $f \times g$  has a domain of definition of the type given in Definition 6.1, it is an associative bilinear transformation. The problem of studying the ways in which Hilbert space can be regarded as a hypercomplex number system is, therefore, the analysis of associative bilinear transformations.

7. We shall show in this section a relationship between the study of as-

sociative bilinear transformations and that of rings of operators of a certain type (cf. [9]). It is customary to assume for a ring of operators  $\mathbf{M}$  that if  $A$  is in  $\mathbf{M}$ ,  $A^*$  is also in  $\mathbf{M}$ . In order therefore to present a precise connection with the theory of rings of operators in its present form, we make the following definition:

DEFINITION 7.1. An associative bilinear transformation  $F(f, g)$  will be said to be closed with respect to adjoints if, whenever  $f$  is in  $\mathfrak{A}$ , there exists an  $f^* \in \mathfrak{A}$  such that  $T_f^* = (T_f)^*$ .

A preliminary connection is the following:

THEOREM 9. The set  $\mathbf{M}$  of  $T_f$ 's associated with a given bilinear transformation  $F(f, g)$  form an algebraic ring of operators<sup>†</sup> for which  $T_f \cdot T_g = T_{F(f, g)}$  if and only if  $F(f, g)$  is associative and closed with respect to adjoints.

Suppose  $F(f, g)$  is associative and closed with respect to adjoints. It follows from Definition 6.1 that each  $T_f$  is bounded. Also if  $A = T_f$  and  $B = T_g$  are in  $\mathbf{M}$ , then  $\alpha A = \alpha T_f = T_{\alpha f}$  is in  $\mathbf{M}$  and  $A + B = T_f + T_g = T_{f+g}$  is in  $\mathbf{M}$ . Since  $F$  is closed with respect to adjoints,  $A^* = (T_f)^* = T_{f^*}$  is in  $\mathbf{M}$ . Finally, since  $f$  and  $g$  are in  $\mathfrak{A}$ ,

$$ABh = T_f \cdot T_g h = F(f, T_g h) = F(f, F(g, h)) = F(F(f, g), h) = T_{F(f, g)} h.$$

Thus  $AB = T_{F(f, g)}$  and  $AB \in \mathbf{M}$ .

Conversely, if the  $T_f$ 's form an algebraic ring in which  $T_f \cdot T_g = T_{F(f, g)}$ , then  $F(f, g)$  must be defined for a linear set of  $f$ 's since the  $T_f$ 's are a linear set and  $T_f + T_g = T_{f+g}$ ,  $\alpha T_f = T_{\alpha f}$ . Since an algebraic ring consists of bounded operators, each  $T_f$  is bounded. Since  $\mathbf{M}$  is multiplicative,  $T_f \cdot T_g = T_{F(f, g)}$  is in  $\mathbf{M}$  and  $F(f, g)$  is in  $\mathfrak{A}$ . Also

$$F(f, F(g, h)) = T_f \cdot T_g h = T_{F(f, g)} h = F(F(f, g), h).$$

Thus  $F$  is associative. Furthermore, since  $A = T_f \in \mathbf{M}$  implies  $A^* \in \mathbf{M}$ ,  $A^* = T_{f^*}$  for some  $f^*$  and  $F$  is therefore closed with respect to adjoints.

Thus the study of  $\times$  operations, subject to the restriction of being closed with respect to adjoints, may be referred to the study of algebraic rings of operators. At present such rings are also subjected in practice to certain further continuity restrictions. We will not need in §§7 and 8 of this paper the full restrictions usually imposed in order to obtain our results, and it is possible that the full restrictions are not needed even in the original theory of rings of operators.

<sup>†</sup> A set of bounded operators  $\mathbf{M}$  forms an algebraic ring if, whenever  $A$  and  $B$  are in  $\mathbf{M}$  and  $\alpha$  is a complex number,  $\alpha A$ ,  $A^*$ ,  $A + B$ , and  $A \cdot B$  are in  $\mathbf{M}$ . This definition is given in [9], p. 383.

**DEFINITION 7.2.** An associative bilinear transformation  $F(f, g)$  will be said to be closed with respect to strongly convergent sequences if, whenever a sequence  $T_{f_n}$  converges strongly to a  $T \in \mathcal{B}$ ,  $T = T_f$  for some  $f \in \mathcal{A}$ .

Closure with respect to strongly convergent sequences is sufficient for us to obtain the essential property of the set of  $T_f$ 's which we need in the following discussion. This property is the existence of a maximal idempotent. For an algebraic ring  $\mathcal{M}$ , the maximal idempotent has been defined in [9], when it exists, as that projection in  $\mathcal{M}$  such that for every  $A \in \mathcal{M}$ ,  $A = AE = EA$ . It is unique.

**THEOREM 10.** If an associative bilinear transformation  $F(f, g)$  is closed with respect to adjoints and strongly convergent sequences, the set  $\mathcal{M}$  of  $T_f$ 's is an algebraic ring closed with respect to strongly convergent sequences. Also  $\mathcal{M}$  contains a maximal idempotent  $E_0 = T_{f_0}$ .

The first statement follows easily from Theorem 9 and Definition 7.2. We must show that an algebraic ring  $\mathcal{M}$ , closed with respect to strongly convergent sequences contains a maximal idempotent.

Now if  $A$  is in  $\mathcal{M}$ , we can find a self-adjoint  $H \in \mathcal{M}$  whose zeros are precisely those of  $A$ . For let  $H = A^*A$ . Then  $Af = 0$  implies  $Hf = 0$ , and  $Hf = 0$  implies  $0 = (A^*Af, f) = \|Af\|^2$ .

The proof given in [9] (II, §2, pp. 389–390) shows that if an algebraic ring  $\mathcal{M}$  is closed with respect to strongly convergent sequences and if  $H$  is in  $\mathcal{M}$ , then  $E(0-)$  and  $1 - E(0)$ , in the resolution of the identity for  $H$ , are each in  $\mathcal{M}$ . (The complete hypothesis that  $\mathcal{M}$  is strongly closed is not used.) Thus  $1 - E(0) + E(0-)$ , the projection on the complement of the zeros of  $H$ , is in  $\mathcal{M}$ .

Combining the results of the two preceding paragraphs we see that for every  $A \in \mathcal{M}$ , we can find a projection  $E \in \mathcal{M}$ , such that the set of zeros of  $E$  is the set of zeros of  $A$ .

We can now show that  $\mathcal{M}$  contains a maximal projection  $E'$ ; that is,  $E'$  is such that for every  $E \in \mathcal{M}$ ,  $E'E = E$ . Now there is a sequence  $\{E_n\}$  strongly dense in the set of  $E \in \mathcal{M}$  (cf. [9], I, §4, pp. 386–388). Since  $E'X$  is continuous in the strong topology, if  $E'E_n = E_n$  for every  $n$ , then  $E'E = E$  for every  $E \in \mathcal{M}$ . Thus it will be sufficient to find an  $E'$  which majorizes the  $E_n$ .

Let  $A_n = E_1 + E_2 + \cdots + E_n$ . Now  $A_nf = 0$  if and only if  $E_if = 0$  for  $i = 1, 2, \dots, n$ , for

$$(A_nf, f) = \sum_{i=1}^n (E_if, f) = \sum_{i=1}^n \|E_if\|^2.$$

Since  $\mathcal{M}$  is linear,  $A_n$  is in  $\mathcal{M}$ . By a preceding result, we can find a projec-

tion  $E'_n$ , whose zeros are precisely those of  $A_n$ . Hence  $E'_n E_i = E_i$  for  $i = 1, 2, \dots, n$ .

We also have, from the definition,  $E'_1 \leq E'_2 \leq \dots$ . It follows that the  $E'_i$  converge strongly to an  $E'$ . Since  $\mathcal{M}$  is closed with respect to strongly convergent sequences,  $E'$  is in  $\mathcal{M}$  and furthermore  $E' E'_n = E'_n$  for every  $n$ . Hence if  $n$  is greater than or equal to  $i$ ,  $E' E_i = E' E'_n E_i = E'_n E_i = E_i$ . Thus  $E'$  majorizes  $E_i$  for every  $i$ , and, as we have remarked above, this is sufficient to yield that  $E'$  is a maximal projection.

But the maximal projection  $E'$  is also a maximal idempotent. For we have  $E' E = E$  for every  $E \in \mathcal{M}$ . Taking adjoints, we also have  $E = E E'$ . Hence  $A = E' A = A E'$  if  $A$  is a projection in  $\mathcal{M}$ . But under these circumstances, this equation must hold for linear combinations of projections too and also for their strong sequential limits. Thus the equation holds for all self-adjoint operators  $A$  in  $\mathcal{M}$ . Finally since an arbitrary bounded  $A$  in  $\mathcal{M}$  is a linear combination of two self-adjoint operators, it holds for every  $A$  in  $\mathcal{M}$ .

Thus  $E_0 = E'$  is a maximal idempotent for  $\mathcal{M}$ . Since it is in  $\mathcal{M}$ , there exists an  $f_0 \in \mathfrak{A}$  such that  $T_{f_0} = E_0$ .

8. In this section, we continue the discussion of  $E_0$ .

LEMMA 8.1. *Let  $F(f, g)$  and  $E_0$  be as above (Theorem 10).  $E_0$  is the projection on the complement of those  $g$ 's such that  $F(f, g) = 0$  for every  $f \in \mathfrak{A}$ .*

**Proof.** If  $g$  is such that  $F(f, g) = 0$  for every  $f \in \mathfrak{A}$ , then  $E_0 g = F(f_0, g) = 0$ . If  $g = (1 - E_0)g$ , then

$$F(f, g) = T_f g = T_f (1 - E_0)g = (T_f - T_f E_0)g = 0 \cdot g = 0.$$

LEMMA 8.2. *If  $f$  is in  $\mathfrak{A}$ ,  $E_0 f$  is in  $\mathfrak{A}$ . If  $f$  is in  $\mathfrak{A}$  and  $(1 - E_0)f = f$ , then  $T_f = 0$ . If  $f$  is in  $\mathfrak{A}$  then  $T_f = T_{E_0 f}$ . If  $\mathfrak{M}$  is the closure of  $\mathfrak{A}$ , then the projection  $E$  on  $\mathfrak{M}$  commutes with  $E_0$ .*

**Proof.** If  $f$  is in  $\mathfrak{A}$ , then  $E_0 f = F(f_0, f)$  is in  $\mathfrak{A}$  by Definition 6.1. If  $f$  is in  $\mathfrak{A}$  and  $(1 - E_0)f = f$ , then

$$T_f = T_{f - E_0 f} = T_f - T_{E_0 f} = T_f - T_{F(f_0, f)} = T_f - E_0 T_f = 0.$$

Also if  $f$  is in  $\mathfrak{A}$ ,

$$T_f = T_{E_0 f} + T_{(1 - E_0)f} = T_{E_0 f}.$$

To show the last statement, we note that  $f = E_0 f + (1 - E_0)f$  with  $E_0 f \in \mathfrak{A}$ ; hence  $(1 - E_0)f \in \mathfrak{A}$ . This implies that the linear set  $\mathfrak{A} = \mathfrak{N}' + \mathfrak{A}'$  where  $\mathfrak{A}'$  is included in the range of  $E_0$  and  $\mathfrak{N}'$  in the orthogonal complement. The closure  $\mathfrak{M}$  is similar, and this implies the last statement of our lemma.

If  $\mathfrak{M}$  is as in Lemma 8.2, we can extend the definition of  $F(f, g)$  so that  $F(f, g)$  is defined for every  $f \in \mathfrak{S} \ominus \mathfrak{M}$ . For if  $f = f_1 + f_2$ ,  $f_1 \in \mathfrak{A}$ ,  $f_2 \in \mathfrak{S} \ominus \mathfrak{M}$ , we may define  $T_f$  as  $T_{f_1}$ . In the resulting extension, the properties of Definition 6.1, 7.1, and 7.2 are preserved (cf. the proof of Corollary 3 below) and, furthermore,  $\mathfrak{A}$  is dense.

**DEFINITION 8.1.** Let  $F(f, g)$  be associative and closed with respect to adjoints and strongly convergent sequences. Let  $\mathfrak{N}_1$  be the closure on the set of  $g$ 's in  $\mathfrak{A}$  such that  $T_g = 0$ . Let  $G_1$  be the projection on  $\mathfrak{N}_1$ . Then  $F(f, g)$  will be said to be regular if, whenever  $f_0$  is such that  $T_{f_0} = E_0$  (cf. Theorem 10),  $G_1 f_0$  is in  $\mathfrak{A}$  and  $T_{G_1 f_0} = 0$ .

Note that if the set of  $g$ 's for which  $T_g = 0$  forms a closed linear manifold, then both conditions are fulfilled and  $F(f, g)$  is regular. It will be shown in this section that regularity implies that there is an extension of  $F$  for which this is the case.

**LEMMA 8.3.**  $G_1$  commutes with  $E_0$ .

**Proof.** If  $g$  is such that  $T_g = 0$ , then  $g = (1 - E_0)g + E_0 g$ . By Lemma 8.2,  $T_{(1-E_0)g} = 0$  and  $T_{E_0 g} = T_g = 0$ . Thus the set of  $g$ 's for which  $T_g = 0$ , is a linear manifold determined by a linear manifold in the range of  $E_0$  and another in the complement of the range of  $E_0$ . The closure  $\mathfrak{N}_1$  has the same property; hence  $G_1$  commutes with  $E_0$ .

**THEOREM 11.** If  $F(f, g)$  is regular, then  $f_0$  (cf. Theorem 10) can be chosen in such a way that

- (a)  $F(f_0, f_0) = f_0$ ,
- (b)  $f_0$  is orthogonal to  $\mathfrak{N}_1$ .

By Lemma 8.2, if  $T_{f_0} = E_0$ , then  $T_{E_0 f_0} = E_0$ . Thus we may choose  $f_0$  so that  $f_0 = E_0 f_0 = T_{f_0} f_0 = F(f_0, f_0)$ .

Now if we let  $f'_0 = f_0 - G_1 f_0$ , where  $E_0 f_0 = f_0$ , then by the regularity of  $F(f, g)$ ,  $f'_0$  is in  $\mathfrak{A}$  and  $T_{f'_0} = T_{f_0} - T_{G_1 f_0} = T_{f_0} = E_0$ . Also by Lemma 8.3,

$$f'_0 = f_0 - G_1 f_0 = E_0 f_0 - G_1 E_0 f_0 = E_0 f_0 - E_0 G_1 f_0 = E_0 (f_0 - G_1 f_0) = E_0 f'_0,$$

and, as we have seen, this implies  $f'_0 = F(f'_0, f'_0)$ . Since  $f'_0 = (1 - G_1)f_0$ ,  $f'_0$  is orthogonal to  $\mathfrak{N}_1$ .

**COROLLARY 1.** Let  $f_0$  be as in Theorem 11, and let  $\mathfrak{A}_0$  be the set of  $f$ 's in  $\mathfrak{A}$  in the form  $F(f, f_0)$ ,  $f \in \mathfrak{A}$ . Then

- (a)  $\mathfrak{A}_0$  is orthogonal to  $\mathfrak{N}$ ;
- (b)  $f \in \mathfrak{A}_0$  and  $T_f f_0 = F(f, f_0) = 0$  imply  $f = 0$  and  $T_f = 0$ ;
- (c) if  $g$  is in  $\mathfrak{A}$ ,  $g = g_0 + g_1$ , where  $g_0$  is in  $\mathfrak{A}_0$  and  $g_1$  is in  $\mathfrak{N}_1$ , and  $T_g = T_{g_0}$ .

This resolution is unique, and  $g_0 = F(g, f_0)$ .

**Proof of (a).** Let  $g$  be such that  $T_g = 0$ . Then

$$(F(f, f_0), g) = (T_f f_0, g) = (f_0, (T_f)^* g) = (f_0, T_f^* g) = (f_0, F(f^*, g)).$$

Now  $T_{F(f^*, g)} = T_f^* \cdot T_g = 0$ . Hence  $F(f^*, g)$  is in  $\mathfrak{N}_1$ . Theorem 11, (b) now yields that

$$0 = (f_0, F(f^*, g)) = (F(f, f_0), g).$$

Thus  $F(f, f_0)$  is orthogonal to  $\mathfrak{N}_1$ . Hence  $\mathfrak{A}_0$  is orthogonal to  $\mathfrak{N}_1$ .

**Proof of (b).** Suppose  $f \in \mathfrak{A}$  and  $F(f, f_0) = 0$ . We note that  $T_{F(f, f_0)} = T_f \cdot T_{f_0} = T_f E_0 = T_f$ . Thus  $F(f, f_0) = 0$  implies  $0 = T_0 = T_{F(f, f_0)} = T_f$ . Thus  $f \in \mathfrak{A}$  and  $f \in \mathfrak{N}_1$ . Hence by (a),  $f = 0$ .

**Proof of (c).** Suppose  $g \in \mathfrak{A}$ . Assume  $g_0 = F(g, f_0) \in \mathfrak{A}$ . Then  $T_g = T_{F(g, f_0)}$ ; hence

$$T_{g_1} = T_{g-g_0} = T_g - T_{g_0} = 0.$$

Since  $\mathfrak{A}_0$  and  $\mathfrak{N}_1$  are orthogonal, the resolution is unique.

**COROLLARY 2.** If  $f_1, f_2, \dots$  are each in  $\mathfrak{A}_0$  (cf. Corollary 1), then

(a)  $F(f_i, f_0) = T_{f_i} f_0 = f_i$ ;

(b) if the sequence  $T_{f_i}$  is strongly convergent with limit  $T$ , then the  $f_i$  converge to an  $f \in \mathfrak{A}_0$ , for which  $T_f = T$ .

**Proof of (a).**  $f_i - F(f_i, f_0)$  is in  $\mathfrak{A}_0$ , since the latter is linear and in  $\mathfrak{N}_1$  by Corollary 1, (c). Hence  $f_i - F(f_i, f_0)$  is in  $\mathfrak{A}_0 \cdot \mathfrak{N}_1$  and is zero by Corollary 1, (a).

**Proof of (b).** Since the  $T_{f_i}$  are convergent,  $f_i = T_{f_i} f_0$  converges to  $T f_0$ . Since  $F$  is closed with respect to strongly convergent sequences (Definition 7.2),  $T = T_g$  for a  $g \in \mathfrak{A}$ . Hence  $f = T f_0 = T_g f_0 = F(g, f_0)$  is in  $\mathfrak{A}_0$ , and by Corollary 1, (c)  $T_f = T_g = T$ .

Corollaries 1 and 2 give the relations for regular  $\times$  operations between the set of  $T_f$ 's and the  $f$ 's in  $\mathfrak{A}$ . Examples will be discussed in §10 below.

We discuss the significance of regularity in the following lemma:

**LEMMA 8.4.** Suppose  $\mathfrak{A}$  is dense. If  $E_1$  is the projection on  $\mathfrak{M}_1$ , the closure of  $\mathfrak{A}_0$ , then  $E_1$  and  $G_1 = 1 - E_1$  each commute with every  $T_f$ .

**Proof.** If  $f$  is in  $\mathfrak{A}_0$ ,  $f = F(f, f_0)$  and

$$T_g f = F(g, f) = F(g, F(f, f_0)) = F(F(g, f), f_0) \in \mathfrak{A}_0.$$

Thus if  $f \in \mathfrak{A}_0$ , then  $T_g f$  is in  $\mathfrak{A}_0$ . Since  $\mathfrak{M}$  is closed with respect to adjoints,  $T_g^* f$  is also in  $\mathfrak{A}_0$ . By continuity, therefore, if  $f \in \mathfrak{M}$ , then  $T_g f \in \mathfrak{M}_1$  and  $(T_g)^* f \in \mathfrak{M}_1$ . Thus  $E_1$  commutes with  $T_f$  (cf. [14], Theorem 4.25).

Since  $\mathfrak{A}$  is dense, it follows from Corollaries 1, (a) and (c) that  $\mathfrak{M}_1$  and  $\mathfrak{N}_1$  are orthogonal complements of each other. Hence  $G_1 = 1 - E_1$  and, of course, commutes with  $T_f$ .

COROLLARY 3. If  $F(f, g)$  is regular,  $F$  has a regular extension in which the set of  $g$ 's for which  $T_g = 0$  form a closed linear manifold.

Suppose first that  $\mathfrak{A}$  is dense. Let  $g$  be in  $\mathfrak{A} \cdot \mathfrak{N}$ . By Corollary 1, (c)  $g = g_0 + g_1$ ,  $g_0 \in \mathfrak{A}_0$ ,  $g_1 \in \mathfrak{N}_1$ . Therefore  $g_0 = g - g_1$  is also in  $\mathfrak{N}_1$  and hence in  $\mathfrak{A}_0 \cdot \mathfrak{N}_1$ . By Corollary 1, (a),  $g_0 = 0$ . Hence  $g = g_1$ ,  $T_g = 0$ .

Thus  $\mathfrak{A} \cdot \mathfrak{N}_1$  consists of those  $g$ 's for which  $T_g = 0$ . Therefore we may extend  $\mathfrak{A}$  to the linear manifold determined by  $\mathfrak{N}_1$  and  $\mathfrak{A}$  as follows. If  $f = f_1 + f_2$ ,  $f_1 \in \mathfrak{A}$ ,  $f_2 \in \mathfrak{N}_1$ , let  $Tf = T_{f_1}$ . This is an extension, since if  $T_f$  is already defined,  $f$  and  $f_1$  are each in  $\mathfrak{A}$ , and thus  $f_2 = f - f_1$  is in  $\mathfrak{A} \cdot \mathfrak{N}_1$  and  $T_{f_2} = 0$ . A similar argument shows that  $Tf$  is unique for  $f$  in the extension of  $\mathfrak{A}$ .

We next show that extension  $F'(f, g)$  is associative (Definition 6.1). Conditions (a) and (b) of that definition are immediately seen to be satisfied. To show (c), let  $f$  and  $g$  be elements of the extension of  $\mathfrak{A}$ . Then  $f = f_1 + f_2$ ,  $f_1 \in \mathfrak{A}$ ,  $f_2 \in \mathfrak{N}_1$  and  $g = g_1 + g_2$ ,  $g_1 \in \mathfrak{A}$ ,  $g_2 \in \mathfrak{N}_1$ . Since  $T_{f_2} = 0$ , we have

$$F'(f_1 + f_2, g_1 + g_2) = F(f_1, g_1 + g_2) = F(f_1, g_1) + F(f_1, g_2) = F(f_1, g_1) + T_{f_2}g_2.$$

Since by Lemma 8.4,  $G_1$  commutes with  $T_{f_1}$ , and since  $g_2$  is in  $\mathfrak{N}_1$ ,  $T_{f_1}g_2$  is in  $\mathfrak{N}_1$ . Also  $F(f_1, g_1)$  is in  $\mathfrak{A}$  and thus  $F'(f_1 + f_2, g_1 + g_2)$  is in the extension of  $\mathfrak{A}$ . To show (d) we note that

$$\begin{aligned} F'(F'(f_1 + f_2, g_1 + g_2), h) &= F'(F(f_1, g_1) + T_{f_2}g_2, h) = F(F(f_1, g_1), h) \\ &= F(f_1, F(g_1, h)) = F'(f_1 + f_2, F'(g_1 + g_2, h)). \end{aligned}$$

Thus  $F'$  is associative.

Definitions 7.1 and 7.2, which are statements concerning the totality of  $T_f$ 's, unaffected in this extension, of course are satisfied by the extension.

In the case in which  $\mathfrak{A}$  is not dense, one first makes the extension given after Lemma 8.2. The  $E$ , which is the projection on  $\mathfrak{M}$ , the closure of  $\mathfrak{A}$ , commutes with every  $T_f$ . For  $g \in \mathfrak{A}$  implies  $T_f g = F(f, g) \in \mathfrak{A}$ . An argument similar to that of Lemma 8.4 then yields that  $E$  commutes with  $T_f$ . The remainder of the justification for this extension is precisely similar to the argument given in the preceding paragraphs. One then makes the further extension given in this proof for the case in which  $\mathfrak{A}$  is dense.

It should also be remarked in connection with the regularity condition, that the restrictions of Definitions 6.1, 7.1, and 7.2 imply little with respect to the set of  $g$ 's for which  $T_g = 0$ . This set may even be dense. The regularity condition removes possibilities of this sort.

9. The previous sections have shown that the analyses of  $\times$  operations leads to an algebraic ring  $\mathbf{M}$  closed with respect to strongly convergent sequences. A further restriction on the nature of  $\mathbf{M}$  is implied in Corollary 1,

(b). This restriction is that there exists an  $f_0$  such that  $Tf_0=0$  implies  $Tf=0$ . If  $M$  is also closed in the strong topology, a certain result ([9], Theorem 5, pp. 393-396) in the theory of rings of operators becomes available, and the nature of this restriction can be explored further.

**THEOREM 12.** *Let  $M$  be a ring of operators (cf. [9], p. 388). Let  $E_0$  be the maximal idempotent of  $M$  (cf. the proof of Theorem 10). Then the necessary and sufficient condition that there should exist an  $f_0$  such that  $T \in M$  and  $Tf_0=0$  imply  $T=0$  is that there should exist an  $f' \in \mathfrak{S}$  such that  $\mathfrak{M}_{E_0f'}^{M'}$  is the range of  $E_0$ .*

**Proof.** Suppose that there is an  $f_0$  such that  $Tf_0=0$  and  $T \in M$  imply  $T=0$ . Since  $T = TE_0$ , this is equivalent to  $TE_0f_0=0$  and  $T \in M$  imply  $T=0$ .

Consider  $\mathfrak{M}_{E_0f_0}^{M'}$ . Since  $E_0 \in M$  if  $A \in M'$ ,  $AE_0f_0 = E_0Af_0$ . Thus  $\mathfrak{M}_{E_0f_0}^{M'}$  is included in the range of  $E_0$ . Furthermore if  $A$  is in  $M'$ ,  $E_{E_0f_0}^{M'}$  commutes with  $A$ .<sup>†</sup> Hence  $E_{E_0f_0}^{M'}$  is in  $M''$ . Since the range of  $E_{E_0f_0}^{M'}$  is included in that of  $E_0$ ,  $E_0E_{E_0f_0}^{M'} = E_{E_0f_0}^{M'}E_0 = E_{E_0f_0}^{M'}$ . The last two statements imply that  $E_{E_0f_0}^{M'}$  is in  $M$  by [9] (Theorem 5, pp. 393-396).

Thus  $E_0 - E_{E_0f_0}^{M'}$  is also in  $M$ . Since 1 is in  $M'$ ,  $E_{E_0f_0}^{M'} \cdot E_0f_0 = E_0f_0$ . Hence

$$(E_0 - E_{E_0f_0}^{M'})E_0f_0 = E_0^2f_0 - E_{E_0f_0}^{M'}E_0f_0 = E_0f_0 - E_0f_0 = 0.$$

Since  $TE_0f_0=0$  and  $T \in M$  imply  $T=0$ , we have  $E - E_{E_0f_0}^{M'} = 0$  or  $E_0 = E_{E_0f_0}^{M'}$ .

Suppose, on the other hand, that there is an  $f'$  such that  $\mathfrak{M}_{E_0f'}^{M'}$  is the range of  $E_0$ . Suppose  $Tf'=0$ ,  $T \in M$ . Then, since  $T = TE_0$ ,  $TE_0f'=0$ . Now if  $A$  is in  $M'$ ,

$$TE_0AE_0f' = ATE_0^2f' = ATE_0f' = A \cdot 0 = 0.$$

Hence  $TE_0$  is zero on the set of  $AE_0f'$ ,  $A \in M'$ . Since this set is dense in the range of  $E_0$  and  $TE_0$  is bounded, this means that  $TE_0$  is zero on the range of  $E_0$ . Since  $TE_0(1-E_0)=0$ , it is zero on the orthogonal complement of this set also. Hence  $TE_0=0$ , and, since  $T = TE_0$ ,  $T=0$ .

**COROLLARY.** *If  $M$  is an algebraic ring closed with respect to strongly convergent series and such that there exists an  $f'$  for which  $\mathfrak{M}_{E_0f'}^{M'} = \mathfrak{M}_0$ , the range of  $E_0$ , then  $f_0 = E_0f'$  is such that  $Tf_0=0$  and  $T \in M$  imply  $T=0$ .*

The last paragraph of the preceding proof shows this.

It may be noted here that there is another way of expressing the condition of Theorem 12 on  $M$ . According to certain unpublished results of J. von

<sup>†</sup> Cf. [7], Definition 5.1.1, p. 143. However in this paper we drop the requirement of this definition that  $M$  should be a prime. This will not affect our use of the known properties of  $\mathfrak{M}_I^M$ .

<sup>‡</sup> Since  $A$  and  $A^*$  are both in  $M'$ , they carry  $\mathfrak{M}_{E_0f_0}^{M'}$  into part of itself. Paper [14], Theorem 4.25, now shows that  $E_{E_0f_0}^{M'}$  commutes with  $A$ .

Neumann, an arbitrary ring  $M$  may be expressed as the "sum" of rings  $M_\alpha$  which are factors on subsets  $\mathfrak{M}_\alpha$  of  $\mathfrak{S}$ , there being a subset for each minimal projection of the center  $M \cdot M'$  and "differential" subsets for the continuous spectrum of  $M \cdot M'$ . If  $M_\alpha$  is considered only on  $\mathfrak{M}_\alpha$ , we may introduce  $M'_\alpha$  within  $\mathfrak{M}_\alpha$ . The condition given above may be restated as follows. For every essential  $\alpha$ , the normalized dimensionality of  $M'_\alpha$  must not be less than that of  $M_\alpha$ .

However, an algebraic ring  $M$  which is subject to the restrictions given in the first paragraph of this section is the set of  $T_f$ 's for an  $\times$  operation of the type which we have considered in this paper.

**DEFINITION 9.1.** Let  $M$  be an algebraic ring of operators (cf. Theorem 9 above) which is closed with respect to strongly convergent series. Furthermore let  $M$  be such that there exists an  $f_0 \in \mathfrak{S}$ , such that  $T \in M$  and  $Tf_0 = 0$  imply  $T = 0$ . We may suppose  $E_0 f_0 = f_0$ .

Let  $\mathfrak{A}_0$  consist of those elements of  $\mathfrak{S}$  in the form  $Tf_0$ ,  $T \in M$ . Let  $\mathfrak{M}_1$  be the closure of  $\mathfrak{A}_0$ ,  $E_1$  the projection of  $\mathfrak{M}_1$ . Let  $\mathfrak{N}_1$  be the orthogonal complement of  $\mathfrak{M}_1$  and  $G_1$  the projection on  $\mathfrak{N}_1$ .

We define  $F_{M,f_0}(f, g)$  (abbreviated to  $F_M(f, g)$ ) as follows. The transformation  $F_M(f, g)$  is defined whenever  $f$  is in the form  $f_1 + f_2$ ,  $f_1 \in \mathfrak{A}_0$ ,  $f_2 \in \mathfrak{N}_1$  (and for all  $g$ ). If, in these circumstances,  $T \in M$  is such that  $f_1 = Tf_0$ , then  $F_M(f, g) = Tg$ .

**THEOREM 13.** Under the assumptions of Definition 9.1,  $F_M(f, g)$  is single-valued. Furthermore  $F_M(f, g)$  is a bilinear transformation, for which the set of  $T_f$ 's is the  $M$  of Definition 9.1.

We show first that  $F_M(f, g)$  is single-valued. For a given  $f$ , there is at most one  $f_1$ , since  $\mathfrak{A}_0$  and  $\mathfrak{N}_1$  are mutually orthogonal. Furthermore there is at most one  $T \in M$  such that  $f_1 = Tf_0$ . For if  $f_1 = Tf_0$  and  $f_1 = T'f_0$ ,  $T$  and  $T' \in M$ , then  $Tf_0 = T'f_0$  and  $(T - T')f_0 = 0$ . The assumptions of Definition 9.1 now imply that  $T - T' = 0$  or  $T = T'$ .

We next show that  $F_M(f, g)$  is a bilinear transformation (Definition 1.1). The previous paragraph shows that for  $f$  fixed, the equation  $T_f g = F(f, g) = Tg$  determines a linear transformation  $T_f$ . If  $g$  is fixed,  $R_g f = F(f, g)$  also defines a linear transformation. For  $R_g f$  is single-valued as we have seen above. Also, inasmuch as  $M$  is linear, if  $f_1 = Tf_0$ ,  $g_1 = Sf_0$ ,  $T$  and  $S \in M$ , then  $af_1 = (aT)f_0$ ,  $f_1 + g_1 = (T + S)f_0$ . Hence

$$F(af, h) = F(a(f_1 + f_2), h) = F(af_1 + af_2, h) = aTh = aF(f_1 + f_2, h) = aF(f, h)$$

by Definition 9.1, and also

$$\begin{aligned} F(f + g, h) &= F(f_1 + g_1 + f_2 + g_2, h) = (T + S)h = Th + Sh \\ &= F(f, h) + F(g, h). \end{aligned}$$

Hence it follows that  $R_\theta$  is linear, and  $F(f, g)$  is a bilinear transformation.

The first paragraph of this proof now shows that the set of  $T_f$ 's for  $F_M(f, g)$  is the  $M$  of Definition 9.1.

**COROLLARY.**  $F_M$  is regular, associative, and closed with respect to adjoints and strongly convergent sequences.

If  $T = T_f, S = T_g$ , then

$$\begin{aligned} F(f_1, g) &= F(f_1 + f_2, g_1 + g_2) = F(f_1, g_1 + g_2) = T_{f_1}(g_1 + g_2) \\ &= T_{f_1}g_1 + T_{f_1}g_2 = F(f_1, g_1) + T_{f_1}g_2. \end{aligned}$$

Now  $E_1 = E_{f_0}^M$  is readily seen to commute with all the  $T_f$  by a familiar argument (cf. the footnote in the proof of Theorem 12). Hence  $G_1 = 1 - E_1$  must also commute with every  $T_f$ , and since  $g_2$  is in  $\mathfrak{R}_1$ ,  $g' = T_{f_1}g_2$  is in  $\mathfrak{R}_1$ . Hence  $T_{g'} = 0$  and  $T_{F_M(f, g)} = T_{F_M(f_1, g_1)} + T_{g'} = T_{F_M(f_1, g_1)}$ . But  $F(f_1, g_1) = T_{f_1}g_1 = T_{g_1} = T \cdot Sf_0 \in \mathfrak{A}_0$ ; hence  $T_{F_M(f_1, g_1)} = T \cdot S$ . Thus  $T_{F_M(f, g)} = T \cdot S = T_f \cdot T_g$ .

Theorems 13 and 9 now imply that  $F$  is associative. Since  $M$  is closed with respect to adjoints and strongly convergent sequences, Definitions 7.1 and 7.2 are satisfied. As we have remarked following Definition 8.1, the fact that the  $g$ 's for which  $T_g = 0$  form a closed linear manifold insures regularity.

10. In this section we wish to discuss briefly known examples of  $\times$  operations.\*

We first refer to a paper of J. von Neumann and the writer [8]. This memoir considers rings of operators  $M$  called "factors in case II<sub>1</sub>," which were discovered in a previous joint paper [7]. These rings, in the case  $\alpha \geq 1$  (cf. [8], §1.1, p. 210), satisfy the assumption of Definition 9.1. For inasmuch as they are closed in the strong topology, they are closed with respect to strongly convergent sequences. Secondly, we may take for  $f_0$ , the uniformly distributed  $g$  of [8] (Theorem II, p. 234). For the maximal idempotent  $E_0$  of these rings is the identity 1, and the  $g$  of [8], Theorem II, shares with the  $f$  of [8], §1.1, the property that  $\mathfrak{M}_f^{M'} = \mathfrak{M}_g^{M'} = \mathfrak{S}$ . Theorem 12, above, now yields that such a  $g$  may be used as the  $f_0$  of Definition 9.1.

If  $\alpha = 1$ , the  $F_M(f, g)$  now resulting from the application of Definition 9.1 has an extension which is the  $\times$  operator of the algebra discussed in [8], chap. 4. The extension is obtained by taking the set of closed operators  $Q_\alpha(M)$ , defined in [8], Definition 4.1.2, as the set of  $T_f$ 's. Since  $Q_\alpha(M)$  contains  $M$ , this is an extension of the  $F_M$  which results from Definition 9.1 above.

If  $\alpha$  is greater than 1, Definition 9.1 is still applicable. The  $\mathfrak{A}_0$  will consist

\* The author is indebted to the referee for the suggestion that the material of this section should be discussed.

of those  $f$ 's in the form  $Tg$ ,  $T \in M$ . Also  $\mathfrak{M}_1$ , the closure of  $\mathfrak{A}_0$ , is  $\mathfrak{M}_1^M$ , and  $E_1 = E_1^M$ . Since  $D_{M'}(\mathfrak{S}) = \alpha > 1 = D_{M'}(E_1)$ , we have a case in which  $E_1$  is less than  $E_0 = 1$ .

The above examples illustrate the case in which  $M$  is abstractly irreducible or, what is the same thing, the case in which  $M$  has a minimal center,  $M \cdot M' = \{\alpha 1\}$ . The opposite extreme is the abelian case, in which  $M \cdot M' = M$ . While we will give later a full discussion of the abelian case (cf. §12), we briefly point out certain simple examples here.

Suppose  $\mathfrak{S}$  is realized as  $\mathfrak{L}_2$ , the set of square summable functions on the interval  $0 \leq x \leq 1$  with

$$(f, g) = \int_0^1 f(x)\bar{g}(x)dx.$$

The set of operators  $M$  defined by the equation  $Tf = \phi(x)f(x)$ , with  $\phi(x)$  bounded, constitutes an abelian ring of operators, as one can readily verify. These operators also satisfy Definition 9.1, since they have the requisite closure property, and we may take  $f_0$  as equal to the element  $f(x) = 1$ . For the resulting  $F_M$ , we have  $F_M(\phi, g) = \phi(x)g(x)$ .

An example which illustrates more completely the considerations of the previous sections is obtained by considering  $\mathfrak{S} \oplus \mathfrak{L}_2 \oplus \mathfrak{L}_2$ , the set of triples  $\{f_1, f_2(x), f_3(x)\}$ . Let  $M$  consist of the operators defined by the equations

$$T\{f_1, f_2(x), f_3(x)\} = \{0, \phi(x)f_2(x), \phi(x)f_3(x)\}$$

in which  $\phi(x)$  is bounded. Let  $f_0 = \{0, 1, 0\}$ . When Definition 9.1 is applied,  $E_0$  is the projection on the set of elements in the form  $\{0, f_2(x), f_3(x)\}$ , and  $\mathfrak{M}_1$ , the range of  $E_1$ , is the set of elements in the form  $\{0, f_2(x), 0\}$ .

In the remainder of this section, we prove that if  $F_M(f, g)$  is everywhere defined and if  $M$  is a ring of operators, then  $M$  is finite dimensional; that is,  $M$  has only a finite number of linearly independent elements. Inasmuch as it is necessary to appeal to the theory of rings of operators, our discussion must be limited to the case in which  $M$  is closed in the strong topology.

But if  $F_M$  is everywhere defined, this restriction is not great. For instance, we can prove the following statement:

**REMARK.** *If  $F_M$  is everywhere defined and is closed (Definition 3.1), then  $M$  is a ring of operators.*

We must show that  $M$  is closed in the strong topology (cf. [9]).

First, we see from Theorem 3 of §3 above that  $F_M$  is continuous.

Secondly, let  $f$  be such that there exists a sequence  $f_n$  in  $\mathfrak{A}$  such that  $T_{f_n}f_0 \rightarrow f$ . Then

$$T_{f_n}f_0 = F(f_n, f_0) = F(f_n, F(f_0, f_0)) = F(F(f_n, f_0), f_0) = F(T_{f_n}f_0, f_0).$$

Since  $F_M(f, g)$  is everywhere defined and continuous,  $f = T_{f_0}f_0$ . Then, since  $F_M$  is continuous, for every  $g$ ,

$$T_{f_n}g = F(f, g) = \lim_{n \rightarrow \infty} F(f_n, g) = \lim_{n \rightarrow \infty} T_{f_n}g.$$

Thirdly, let  $T$  be a limit point of  $M$  in the strong topology. We show that  $T$  is in  $M$ ; that is,  $T = T_f$  for some  $f$ . For, let  $f = T_{f_0}$ , and let  $g$  be any element of  $\mathfrak{H}$ . Then since  $T$  is a strong limit of  $M$ , we can find a sequence of  $f_n$ 's such that  $T_{f_n}f_0 \rightarrow T_{f_0} = f$  and  $T_{f_n}g \rightarrow Tg$ . Hence, by the above,  $Tg = \lim_{n \rightarrow \infty} T_{f_n}g = Tg$ . Thus  $T = T_f$ , and  $M$  is closed in the strong topology.

LEMMA 10.1. *Let  $M$  be a ring of operators (cf. [9], p. 388). Furthermore let  $M$  be such that there exists an  $f_0 \in \mathfrak{H}$ , such that  $T \in M$  and  $Tf_0 = 0$  imply  $T = 0$ . Let  $A$  be an abelian ring in  $M$ . Let  $\mathfrak{B}_0$  consist of those elements in the form  $Af_0$ ,  $A \in A$ . Let  $[\mathfrak{B}_0]$  denote the closure of  $\mathfrak{B}_0$ . Then  $[\mathfrak{B}_0] \cdot \mathfrak{A}_0 = \mathfrak{B}_0$ .*

Let  $f$  be in  $\mathfrak{A}_0 \cdot [\mathfrak{B}_0]$ . Then  $f = Tf_0$ ,  $T \in M$ . Since  $f$  is in  $[\mathfrak{B}_0]$ , we can find a sequence  $\{A_n\}$ ,  $A_n \in A$ , such that  $A_nf_0 \rightarrow f$  or  $A_nf_0 \rightarrow Tf_0$ .

Let  $\mathfrak{K} \subset \mathfrak{H}$  consist of those  $g$ 's such that  $g = Bf_0$ ,  $B \in M'$ . Then  $A_ng = A_nBf_0 = BA_nf_0 \rightarrow BTf_0 = TBf_0 = Tg$ . Thus  $A_ng \rightarrow Tg$  for all  $g \in \mathfrak{K}$ .

Theorem 13 above states that  $\mathfrak{K}$  is dense in the range of  $E_0$ . Since  $A_nE_0 = A_n$  and  $TE_0 = T$ ,  $A_ng = Tg = 0$ , for  $g$  in the complement of the range of  $E_0$ . It follows that  $A_ng \rightarrow Tg$  for a dense set of  $g$ 's.

However, it can also be shown that the  $A_n$  can be chosen so that they converge to a closed  $A \cap A$ .<sup>\*</sup> We present here merely an outline of the proof of this statement. The omitted details can easily be seen if use is made of the consideration of [14], chaps. 6 and 7. It is a consequence of [9] (III, §2, pp. 401-404) that there exist a resolution of the identity  $E(\lambda)$  and bounded functions  $\phi_n(\lambda)$  such that  $A_n = \int_0^1 \phi_n(\lambda) dE(\lambda)$ . Since  $Af_0 = 0$  and  $A \in A$  imply  $A = 0$ , it follows that for  $f \in \mathfrak{M}_{f_0}^A$ , the equation  $f = \int_0^1 \phi(\lambda) dE(\lambda)f_0$  determines  $\phi$  essentially.<sup>†</sup> If

$$f = Tf_0 = \int_0^1 \phi(\lambda) dE(\lambda)f_0,$$

then, since  $A_nf_0 \rightarrow f$ , it follows that the  $\phi_n(\lambda)$  approach  $\phi$  essentially; that is,

$$\lim_{n \rightarrow \infty} \int_0^1 |\phi - \phi_n|^2 d\mu = 0.$$

<sup>\*</sup> The statement  $A \cap A$ , means that  $A$  has a dense domain, commutes with all unitary operators in  $A'$ , and is zero on the set on which all  $A \in A$  is zero. This is a modification of [8], Definition 4.2.1.

<sup>†</sup> If  $\mu(\lambda) = \|E(\lambda)f_0\|^2$ , the word "essentially" refers to this  $\mu$ . For instance,  $f$  determines  $\phi$  except possibly on a set of  $\mu$ -measure zero.

Now the  $A_n$ 's were any sequence in  $A$  such that  $A_n f_0 \rightarrow f$ . It is now clear that the  $A_n$ 's could be chosen in such a way that  $|\phi_n(\lambda)|$  is an increasing sequence for each  $\lambda$ . These results are sufficient to show that if  $A = \int_0^1 \phi(\lambda) dE(\lambda)$ , then  $A_n g \rightarrow Ag$  for every  $g$  in the domain of  $A$  and that the sequence  $A_n$  converges only on this domain.

Thus  $Ag = Tg$  for all  $g$  in a dense linear set. But if  $T'$  is the contraction of  $T$ , whose domain is this set,  $T'$  has a unique closed extension, which is bounded. Thus  $A = T$  and is bounded. Hence if  $f = Tf_0$  is in  $[\mathfrak{B}_0] \cdot \mathfrak{A}_0$ , then  $f = Af_0$ ,  $A \in A$ , or  $f$  is in  $\mathfrak{B}_0$ . So  $\mathfrak{B}_0 \supset [\mathfrak{B}_0] \cdot \mathfrak{A}_0$ . But since obviously  $\mathfrak{B}_0 \subset [\mathfrak{B}_0] \cdot \mathfrak{A}_0$ , we have  $\mathfrak{B}_0 = [\mathfrak{B}_0] \cdot \mathfrak{A}_0$ .

LEMMA 10.2. Let  $M$  be as in Lemma 10.1; then if  $M$  contains an infinite abelian ring  $A$ ,  $F_M(f, g)$  is not everywhere defined.

**Proof.** As we pointed out in the proof of Lemma 10.1, each element of  $A$  is in the form  $\int_0^1 \phi(\lambda) dE(\lambda)$  for a fixed resolution of the identity  $E(\lambda)$ . Furthermore  $\phi(\lambda)$  is bounded. However, it is also easily seen that for every  $\phi(\lambda)$  such that  $\int_0^1 |\phi(\lambda)| d\|E(\lambda)f_0\|^2$  exists, there exists an  $f \in \mathfrak{F}$  such that  $f = \int_0^1 \phi(\lambda) dE(\lambda)f_0$ .

As we remarked above, if  $f$  is given,  $\phi(\lambda)$  is essentially unique.

We next exhibit an unbounded  $\phi$  which is such that

$$\int_0^1 |\phi(\lambda)|^2 d\|E(\lambda)f_0\|^2 < \infty.$$

Now inasmuch as  $A$  is infinite, we can divide the interval  $(0, 1)$  into a denumerably infinite number of mutually exclusive subintervals  $I_n$  with adjoints  $a_n$  and  $b_n$  such that if  $E(I_n) = E(b_n) - E(a_n)$ , then  $E(I_n) \neq 0$  and  $E(I_n) \cdot E(I_m) = 0$  if  $n \neq m$ . Now  $E(I_n)f_0 \neq 0$ , for  $E(I_n)f_0 = 0$  implies  $E(I_n) = 0$ , since  $E(I_n)$  is in  $M$ . Also

$$\sum_{n=1}^{\infty} \|E(I_n)f_0\|^2 = \|f_0\|^2.$$

Let now  $\alpha_n = \|E(I_n)f_0\|^2$ . Then  $\alpha_1 + \alpha_2 + \alpha_3 + \dots$  is a convergent sequence of positive terms. It is a well known result in the theory of infinite series, that we can find another convergent series of positive terms,

$$\lambda_1 \alpha_1 + \lambda_2 \alpha_2 + \lambda_3 \alpha_3 + \dots$$

such that  $\lim_{n \rightarrow \infty} \lambda_n = \infty$ .<sup>\*</sup> Now for  $x \in I_n$ , let  $\phi(x) = (\lambda_n)^{1/2}$ . The function  $\phi(x)$  is unbounded and

<sup>\*</sup> Let  $R_n$  denote the remainder in the  $\alpha$  series after the  $n$ th term. Let  $n_k$  denote that number such that for  $n \geq n_k$ ,  $2^k R_n \leq 1/2^k$ . Now if  $n$  is such that  $n_{k-1} < n \leq n_k$ , let  $\lambda_n = 2^{k-1}$ . Then  $\lambda_1 \alpha_1 + \lambda_2 \alpha_2 + \dots = R_0 - R_{n_1} + 2(R_{n_1} - R_{n_2}) + 2^2(R_{n_2} - R_{n_3}) + \dots \leq R_0 + 2R_{n_1} + 2^2 R_{n_2} + \dots \leq R_0 + 1/2 + 1/2^2 + \dots = R_0 + 1$ .

$$\int_0^1 |\phi(\lambda)|^2 d\|E(\lambda)f_0\|^2 = \lambda_1\alpha_1 + \lambda_2\alpha_2 + \cdots < \infty.$$

Now if  $f = \int_0^1 \phi(\lambda) dE(\lambda)f_0$ , then  $f$  is not in  $\mathfrak{B}_0$ . For since  $f$  determines  $\phi$  essentially,  $f \in \mathfrak{B}_0$  implies that  $\phi$  is bounded. But  $f$  is easily seen to be in  $[\mathfrak{B}_0]$ . For if we define  $\phi_n$  as equal to  $\phi(x)$ , when  $\phi(x)$  is less than or equal to  $n$  and equal to  $n$  otherwise, then  $A_n = \int_0^1 \phi_n(\lambda) dE(\lambda)$  is in  $\mathcal{A}$  and  $f_n = A_n f_0$  is in  $\mathfrak{B}_0$ . Furthermore  $f = \lim_{n \rightarrow \infty} f_n$ ; so  $f$  is in  $[\mathfrak{B}_0]$ .

Now  $f$  is not in  $\mathfrak{A}_0$ . For  $f \in \mathfrak{A}_0$  implies  $f \in [\mathfrak{B}_0] \cdot \mathfrak{A}_0 = \mathfrak{B}_0$  (Lemma 10.1). Since  $f$  is in the closure of  $\mathfrak{A}_0$  but not in  $\mathfrak{A}_0$ ,  $f$  is not in  $\mathfrak{A}$ .

Thus  $F_M(f, g)$  is not defined.

The statement of the lemma is still valid even if we permit  $M$  to include closed unbounded operators which are limits of transformations in  $M$  on their domain but preserve, for the enlarged set, the property that  $A \in M$  and  $Af_0 = 0$  imply  $A = 0$ . For these the  $f$  in the proof of Lemma 10.2 is such that  $T_f = \int_0^1 \phi(\lambda) dE(\lambda)$  since  $T_f$  is unique. But since  $T_f$  is unbounded, its domain is not the full space, and there are  $g$ 's for which  $T_f g = F_M(f, g)$  is not defined.

LEMMA 10.3. *If  $M$  is infinite, it contains an infinite abelian ring.*

Suppose  $M$  contains only finite abelian rings. In particular, then the center  $M \cdot M' = A$  is finite. This is easily seen to mean that there exists a finite number of mutually orthogonal projections,  $E_1, E_2, \dots, E_n$ , such that  $A$  is the set of transformations in the form  $\sum_{i=1}^n a_i E_i$ . We also know that  $E_0$  is in  $M \cdot M' = A$  and  $E_0 E_i = E_i$  for  $i = 1, 2, \dots, n$ . Hence  $E_0 = \sum_{i=1}^n E_i$ .

Thus if  $A$  is in  $M$ ,

$$A = A E_0 = A \left( \sum_{i=1}^n E_i \right) = \sum_{i=1}^n A E_i = \sum_{i=1}^n E_i A E_i.$$

Let  $\mathfrak{M}_a$  denote the range of  $E_a$ , and let  $M_a$  denote the set of transformations in the form  $A_a = E_a A E_a$ ,  $A \in M$ , considered only on  $\mathfrak{M}_a$ . The set  $M_a$  contains no elements  $A_a$  except those in the form  $a \cdot 1$  which commute with every  $A_a$ . For if  $A_a$  is such,  $E_a A_a E_a$  is in  $M$ , and also, as we see from the form of an arbitrary  $A \in M$ , it commutes with every  $A \in M$ . Hence  $E_a A_a E_a$  is in  $M \cdot M' = A$ . This implies that  $E_a A_a E_a$  is in the form  $\sum_{i=1}^n a_i E_i$ . Hence  $E_a A_a E_a = a_a E_a$ . Thus  $A_a = a \cdot 1$  on  $\mathfrak{M}_a$  and  $M_a \cdot M'_a = (a \cdot 1)$ .

Furthermore  $M_a$  does not contain an infinite abelian ring  $A_a$ . For if it did, the abelian ring  $A_a$  consisting of transformations in the form  $E_a A_a E_a$ ,  $A_a \in A_a$ , would be an infinite abelian ring in  $M$ .

An analysis of rings for which  $M \cdot M' = \{a \cdot 1\}$  is found in [7]. There are five types of such rings, cases I<sub>n</sub>, I<sub>∞</sub>, II<sub>1</sub>, II<sub>∞</sub>, III<sub>∞</sub> (cf. [7], Theorem VIII, p. 172). It is a characteristic of cases II<sub>1</sub>, II<sub>∞</sub>, III<sub>∞</sub>, that they do not contain a

minimal projection (cf. [7], Definition 5.1.2). Hence they must contain an infinite abelian ring. Since  $I_\infty$  is isomorphic to the set of all operators on Hilbert space, it too contains infinite abelian rings. Hence since  $M_\alpha$  is such that  $M_\alpha \cdot M_{\alpha'} = (a \cdot 1)$  and does not contain an infinite abelian ring,  $M_\alpha$  must be an  $I_n$ .

This, of course, means that  $M_\alpha$  is a finite dimensional ring. For  $A \in M$ , we know that  $A = \sum_{i=1}^n E_\alpha A_\alpha E_\alpha$ ,  $A_\alpha \in M_\alpha$ ; so  $M$  itself is finite dimensional. We have thus shown that if  $M$  contains only finite abelian rings, it is finite.

**THEOREM 14.** *If  $F(f, g)$  is a regular associative bilinear transformation such that the set of  $T_i$ 's forms a ring of operators (that is, is closed with respect to adjoints and also closed in the strong topology), then if  $F(f, g)$  is everywhere defined,  $M$  is finite dimensional.*

It is a consequence of the discussion of §8, that if  $M$  is the set of  $T_i$ 's for  $F(f, g)$ , then  $F(f, g) = F_M(f, g)$  (Definition 9.1). Since  $F(f, g)$  is everywhere defined, we see from Lemma 10.2, that  $M$  does not contain an infinite abelian ring, and hence by Lemma 10.3 that it is finite dimensional.

The remark at the beginning of this discussion (preceding Lemma 10.1) now shows that the following statement is true:

**COROLLARY.** *If  $F$  is a regular associative bilinear transformation which is closed with respect to adjoints and closed (Definition 3.1) and if  $F$  is everywhere defined, then  $M$ , the set of  $T_i$ 's, is finite dimensional.*

11. If  $M$  is a ring of operators for which Definition 9.1 is applicable and if  $F_M(f, g)$  is an associated bilinear transformation, then the  $R_\alpha$ 's of  $F_M$  have certain properties which we discuss in this section. We let  $\mathfrak{A}_0$ ,  $\mathfrak{M}_1$ ,  $E_1$ ,  $\mathfrak{N}_1$ , and  $G_1$  be as in Definition 9.1. In conformity with §§7 and 8, we let  $E_0$  be the maximal idempotent in  $M$  (cf. Theorem 10),  $\mathfrak{M}_0$  the range of  $E_0$ , and  $\mathfrak{A}$  the set of elements in the form  $f_1 + f_2$ ,  $f_1 \in \mathfrak{A}_0$ ,  $f_2 \in \mathfrak{N}_1$ .

There are certain relations in this situation which will be used without further comment. Thus  $\mathfrak{A}$  is dense by Definition 9.1. Lemma 8.4 yields  $E_1 = 1 - G_1$ . The second sentence of Lemma 8.2 implies  $1 - E_0 \leq G_1$ . These two relations yield further  $E_1 \leq E_0$ .

**THEOREM 15.** *Let  $M$  be a ring of operators (cf. [9], p. 388) such that Definition 9.1 is applicable. Let  $F_M(f, g)$ ,  $E_1$ ,  $E_0$ , and  $\mathfrak{A}$  be as in the preceding two paragraphs. Then*

(a)  $R_\alpha f = F_M(f, g)$  is a linear transformation whose domain is  $\mathfrak{A}$  and  $R_\alpha \eta \in M'$  (cf. [7], Definition 4.2.1, p. 141);

(b) the set of  $[R_\alpha]$  for  $R_\alpha$  bounded is exactly the set of  $R \in M'$  for which  $E_0 R = R = R E_1$ .

**Proof of (a).** By the bilinearity of  $F_M(f, g)$ ,  $R_\theta$  is a linear transformation. The domain of  $R_\theta$  is the set of  $f$ 's for which  $F_M(f, g)$  is defined, and Definition 9.1 shows that this set is  $\mathfrak{A}$ . We must show  $R_\theta \eta M'$ . Now if  $A$  is in  $M'$ ,

$$A = A_1 + (1 - E_0)B(1 - E_0), \quad A_1 \in M,$$

and  $B$  can be quite arbitrary (cf. [9], Theorem 5, p. 393). Now since  $A_1$  is in  $M$ , if  $f_1 = A_1 f_0$ , then for  $g$  arbitrary,  $F_M(f_1, g) = A_1 g$ . Also if  $h$  is in the domain of  $R_\theta$ , then

$$(1 - E_0)R_\theta h = (1 - E_0)F(h, g) = (1 - E_0)T_h g = (1 - E_0)E_0 T_h g = 0.$$

Hence  $(1 - E_0) \cdot R_\theta = 0$ . Furthermore, for  $h$  arbitrary,  $(1 - E_0)h$  is in  $\mathfrak{A}_1$  and hence in  $\mathfrak{A}$ , the domain of  $R_\theta$ . Also  $T_{(1-E_0)h} = 0$ . Thus

$$R_\theta(1 - E_0)h = F((1 - E_0)h, g) = T_{(1-E_0)h} g = 0;$$

so  $R_\theta(1 - E_0) = 0$ .

Thus for every  $h$  in the domain of  $R_\theta$ ,

$$\begin{aligned} AR_\theta h &= (A_1 + (1 - E_0)B(1 - E_0))R_\theta h = A_1 R_\theta h = A_1 F(h, g) \\ &= F(f_1, F(h, g)) = F(F(f_1, h), g) = R_\theta F(f_1, h) = R_\theta A_1 h \\ &= R_\theta(A_1 + (1 - E_0)B(1 - E_0))h = R_\theta A h. \end{aligned}$$

Hence  $AR_\theta \subset R_\theta A$ . Since this is true also for  $A^*$ , we see that  $R_\theta \eta M'$ .

**Proof of (b).** Since  $E_1$  is the projection on the closure of  $\mathfrak{A}_0$ ,  $1 - E_1$  is the projection on the zero's of  $R_\theta$ ; hence  $R_\theta(1 - E_1) = 0$  or  $R_\theta = R_\theta E_1$ . From (a) above, we see that  $R_\theta E_1 = R_\theta \eta M'$ . Now  $R_\theta$  has domain dense. Hence if  $R_\theta = R_\theta E_1$  is bounded, it has a closed extension  $[R_\theta] \in M'$ . Furthermore,

$$[R_\theta] = [R_\theta E_1] = [R_\theta E_1] \cdot E_1 = [R_\theta] E_1 = E_0 [R_\theta].$$

Let  $R$  be in  $M'$  and such that  $RE_1 = R = E_0 R$ . Let  $g = Rf_0$ . Then  $g = E_0 Rf_0$  or  $g$  is in the range of  $E_0$ . Hence  $F_M(E_0 f_0, g) = E_0 g = g$ . Consider  $R_\theta$ . For  $f \in \mathfrak{A}$ , we have

$$\begin{aligned} R_\theta f &= R_\theta T_f f_0 = R_\theta T_f E_0 f = T_f R_\theta E_0 f = T_f F_M(E_0 f, g) \\ &= T_f E_0 g = T_f g = T_f Rf_0 = RT_f f_0 = Rf, \end{aligned}$$

using the fact that  $R_\theta$  commutes with  $T_f$  and that  $T_f = T_f E_0$ . Hence  $R$  is an extension of  $R_\theta$  and  $R_\theta$  is bounded. Since the domain of  $R_\theta$  is dense,  $R = [R_\theta]$ . Thus the set of  $[R_\theta]$  includes the set of elements of  $M'$ , for which  $R = RE_1 = E_0 R$ . The results of the previous paragraph now show that these two sets are equal.

**COROLLARY 1.**  $R_\theta$  is bounded for a dense linear set  $\mathfrak{D}$  of  $g$ 's.

Let us denote by  $\mathfrak{D}_0$ , the set of  $g$ 's in the form  $Af_0$ ,  $A \in \mathbf{M}'$ . Since  $\mathbf{M}$  is a ring, Theorem 12 now states that  $\mathfrak{D}_0$  is dense in the range of  $E_0$  and  $Af_0 = E_0Af_0$ . But  $E_0$  is in  $\mathbf{M}'$ , and Lemma 8.4 implies that  $E_1$  is in  $\mathbf{M}'$ . Thus if  $A$  is in  $\mathbf{M}'$ ,  $E_0AE_1$  is in  $\mathbf{M}'$ . Also since  $f_0$  is in  $\mathfrak{A}$ ,  $E_1f_0 = f_0$ . Hence  $E_0AE_1f_0 = E_0Af_0 = Af_0$ . So if  $g$  is in  $\mathfrak{D}_0$  and  $g = Af_0$ ,  $A \in \mathbf{M}'$ , we can suppose that  $A = E_0A = AE_1$ .

Theorem 15, (b) now implies that  $A = [R_g]$ ; so  $R_g$  is bounded for  $g \in \mathfrak{D}_0$ . Furthermore, if  $g$  is in the range of  $1 - E_0$  and  $f \in \mathfrak{A}$ , then

$$R_gf = F(f, g) = Tfg = T(1 - E_0)g = TE_0(1 - E_0)g = 0.$$

Since  $\mathfrak{A}$  is dense,  $[R_g] = 0$ . Thus  $R_g$  is bounded if  $g = g_1 + g_2$ ,  $g_1 \in \mathfrak{D}_0$ ,  $g_2$  in the range of  $1 - E_0$ . Denote the set of such  $g$ 's by  $\mathfrak{D}$ . Since  $\mathfrak{D}_0$  is dense in the range of  $E_0$ ,  $\mathfrak{D}$  is dense.

For  $g \in \mathfrak{D}$ , we may suppose that  $F_M(f, g)$  is defined for all  $f$ . For this extension  $F_M'(f, g)$  we have the following corollaries:

**COROLLARY 2.** *If  $\mathfrak{D}$  is as in the proof of Corollary 1, and  $f$  and  $g$  are in  $\mathfrak{D}$ , then  $F_M'(f, g)$  is in  $\mathfrak{D}$ .*

Let  $f = f_1 + f_2$ ,  $g = g_1 + g_2$ ,  $f_1$  and  $g_1 \in \mathfrak{D}_0$ ,  $f_2$  and  $g_2$  in the range of  $1 - E_0$ . By Lemma 8.2,  $Tf_2 = 0$ . Since we also have  $Tfg = T_f(1 - E_0)g = 0$ ,

$$F_M(f, g) = F_M(f, g_1 + g_2) = F_M(f, g_1) = F_M(f_1 + f_2, g_1) = F_M(f_1, g_1).$$

Let  $R_{f_1} = A$ ,  $R_{g_1} = B$ . Then

$$F_M(f, g) = F_M(f_1, g_1) = Bf_1 = BAf_0.$$

Since  $A$  and  $B$  are in  $\mathbf{M}'$ ,  $BA$  is in  $\mathbf{M}'$ . Also

$$BA = (E_0B) \cdot A = E_0(BA), \quad BA = B(AE_1) = (BA)E_1.$$

Thus  $BA$  is an  $R_A$  by Theorem 15, (b). Furthermore, the proof of Theorem 15, (b) shows that since  $F_M(f, g) = BAf_0$ ,  $BA = R_{F_M(f, g)}$ . Hence  $F_M(f, g)$  is in  $\mathfrak{D}_0$ , which is included in  $\mathfrak{D}$ .

**COROLLARY 3.** *The set of  $R_g$ 's,  $g \in \mathfrak{D}_0$ , is closed in the strong topology.*

**Proof.**  $\mathbf{M}'$  is closed in the strong topology; hence the set of  $A$ 's of  $\mathbf{M}'$  for which  $A = AE_1 = E_0A$  is also closed in the strong topology. Theorem 15, (b) now yields the corollary.

Suppose, for the moment, that Definition 6.1 had been defined with respect to the second variable of  $F_M(f, g)$  rather than the first. Corollary 1 is then the statement that conditions (a) and (b) are fulfilled by the extension of  $F_M$ , and Corollary 2 is the same with respect to (c). Condition (d) is symmetric in the two variables.  $F_M$  thus satisfies Definition 6.1 in the new form.

Corollary 3 carries the topological closure property of the  $T_f$ 's over to the  $R_g$ 's. We next discuss closure with respect to adjoints for the  $R_g$ 's.

**COROLLARY 4.** *The set of  $R_g$ 's is such that for every  $g \in \mathfrak{D}$  there exists a  $g^*$  such that  $[R_g]^* = [R_{g^*}]$ , if and only if  $E_0 = E_1$ .*

Suppose that for every  $g \in \mathfrak{D}$  there exists a  $g^*$  such that  $[R_g]^* = [R_{g^*}]$ . Then, since for a closed bounded  $T$  we have  $T^{**} = T$ , every  $[R_g]$  is the adjoint of a transformation  $[R_{g^*}]$ . Now

$$[R_g]^* = [R_g \cdot E_1]^* = ([R_g \cdot E_1] \cdot E_1)^* = E_1 [R_g \cdot E_1]^* = E_1 [R_g]^*.$$

Since  $[R_g] = [R_{g^*}]^*$ , we have  $[R_g] = E_1 [R_g]$ .

In the first paragraph of the proof of Corollary 1 of this section, it is shown that if, for an element  $f$ ,  $f = Af_0$  with  $A$  in  $\mathbf{M}$ , then  $A$  may be chosen so that  $A = E_0 A = A E_1$ . Hence by Theorem 15, (b),  $A = [R_g]$  for some  $g$ . So if  $f = Af_0$ ,  $A \in \mathbf{M}'$ , then  $f = [R_g]f_0$ . Hence

$$\begin{aligned} \mathfrak{M}_{E_0 f_0}^{\mathbf{M}'} &= [\{Af_0, A \in \mathbf{M}'\}] = [\{[R_g]f_0, g \in \mathfrak{D}\}] = [\{E_1[R_g]f_0, g \in \mathfrak{D}\}] \\ &= E_1[\{[R_g]f_0, g \in \mathfrak{D}\}] = E_1 \mathfrak{M}_{E_0 f_0}^{\mathbf{M}'} \end{aligned}$$

Theorem 12 states that  $\mathfrak{M}_{E_0 f_0}^{\mathbf{M}'}$  is the range of  $E_0$ . Thus  $E_1 \mathfrak{M}_{E_0 f_0}^{\mathbf{M}'} = \mathfrak{M}_{E_0 f_0}^{\mathbf{M}'}$  implies  $E_1 \geq E_0$ . Since we also have  $E_1 \leq E_0$ , we have  $E_1 = E_0$ . Thus if for every  $g \in \mathfrak{D}$  there exists a  $g^* \in \mathfrak{D}$  such that  $[R_g]^* = [R_{g^*}]$ , then  $E_1 = E_0$ .

The converse of this result is given immediately by Theorem 15, (b).

The results of this section show that for regular associative bilinear transformations,  $F_M(f, g)$ , the properties of the two variables are the same if and only if  $E_1 = E_0$ . However if  $E_1 = E_0$  this symmetry extends even further since the ring consisting of transformations in the form  $E_1 A E_1$ ,  $A \in \mathbf{M}'$ , is related to the second variable as  $\mathbf{M}$  is to the first in Definition 9.1.

12. In the special case of a regular associative bilinear transformation in which the  $T_f$  commute, that is,  $F(f, g) = F(g, f)$ ,  $f$  and  $g \in \mathfrak{A}_0$ , closure for strongly convergent sequences is equivalent to strong closure (cf. [9], III, §1, p. 398). The known analyses of self-adjoint operators and abelian rings then permit us to obtain more specific results.

**THEOREM 16.** *Let  $F$  be a regular associative bilinear transformation closed with respect to strongly convergent sequences and adjoints. Furthermore, let  $F(f, g) = F(g, f)$  for  $f$  and  $g \in \mathfrak{A}_0$  (cf. Corollary 1 of Theorem 11). Let  $E_0$ ,  $E_1$ , and  $f_0$  be as in §§7 and 8. Then*

(a)  $\mathbf{M}$  is generated by a self-adjoint transformation

$$H = \int_0^{\epsilon} \lambda dE(\lambda);$$

(b) if  $\mu(\lambda) = \|E(\lambda)f_0\|^2$ , there exists a sequence, finite or infinite, of  $\mu$ -measurable sets,  $S_1, S_2, S_3, \dots$ ,  $S_1 \supseteq S_2 \supseteq S_3 \supseteq \dots$ , such that  $\mathfrak{S}$  may be realized as  $\mathfrak{S}_0 \oplus \mathfrak{S}_1 \oplus \mathfrak{S}_2 \oplus \dots$  (cf. [14], Theorems 1.26 and 1.27), where  $\mathfrak{S}_0$  corresponds to the elements in the range of  $1 - E_0$  and  $\mathfrak{S}_i$  is the space of  $\mu$ -summable squared functions  $\psi_i(\lambda)$  defined on  $S_i$ ;

(c)  $E_1$  is the projection on the set of elements in the form  $\{0, \phi_1(\lambda), 0, \dots\}$ ,  $\mathfrak{A}_0$  the subset of these for which  $\phi_1(\lambda)$  is essentially bounded,  $f_0$  that one of these elements for which  $\phi_1(\lambda) = 1$ ;

(d) the relations

$$F(\{0, \phi_1(\lambda), 0, \dots\}, \{g_1, \psi_1(\lambda), \psi_2(\lambda), \dots\}) = \{0, \phi_1(\lambda)\psi_1(\lambda), \phi_1(\lambda)\psi_2(\lambda), \dots\},$$

and

$$T_{\{0, \phi_1(\lambda), \dots\}} = \int_0^1 \phi_1(\lambda) dE(\lambda)$$

are satisfied;

(e) if  $f = \{0, \phi_1(\lambda), 0, \dots\}$  is in  $\mathfrak{A}_0$ , then  $f^* = \{0, \bar{\eta}_1(\lambda), 0, \dots\}$ .

**Proof.**  $M$  is, as we have seen above, an abelian ring, and this (cf. [9], III, §2, pp. 401-404) implies (a). We now apply the analysis of [14] (chap. 7, §2) to  $H$  considered only on the range of  $E_0$ . We do not, however, distinguish between the point and continuous spectrum, the point spectrum representing merely discontinuities of the  $\rho_i(\lambda)$ . Since for  $E \in M$ ,  $Ef_0 = 0$  implies  $E = 0$ , we may take  $\rho_1(\lambda) = \mu(\lambda)$ . We obtain a sequence of functions of bounded variation  $\rho_1 \prec \rho_2 \prec \rho_3 \prec \dots$ , such that we can express the range of  $E_0$  in the form  $\mathfrak{S}_1 \oplus \mathfrak{S}_2 \oplus \mathfrak{S}_3 \oplus \dots$  where  $\mathfrak{S}_i$  is the space whose inner product is

$$\int_0^e \phi(\lambda) \bar{\theta}(\lambda) d\rho_i(\lambda) = \int_0^e \phi(\lambda) \bar{\theta}(\lambda) \frac{d\rho_i}{d\rho_1} d\rho_1.$$

Now if  $S_i$  is the set on which  $d\rho_i/d\rho_1 \neq 0$  and if, to the element  $\phi(\lambda)$  in this realization, we make  $\phi' = \phi(\lambda)(d\rho_i/d\rho_1)^{1/2}$  correspond, we see that  $\mathfrak{S}_i$  may also be realized as the space whose inner product is  $\int_{S_i} \phi'(\lambda) \bar{\theta}'(\lambda) d\rho_1$ . We have essentially, except for the sets of  $\mu$ -measure zero,  $S_1 \supset S_2 \supset S_3 \supset \dots$ , and this completes the proof of (b).

The above process has already identified  $\mathfrak{S}_1$  with  $\mathfrak{M}_1$ , and the remaining statements are immediate consequences of the operational calculus (cf. [14], chap. 6, or [10]).

**Examples.** We present here five examples.

**EXAMPLE 1.** We give first an example of a bilinear transformation whose domain is completely linear but not rectangular. Let  $\{\phi_i\}$  be an orthonormal set containing at least two elements. Let  $F(k\phi_i, l\phi_i) = kl\phi_i$  for each  $i$ . As to

when  $F(f, 0)$  and  $F(0, g)$  are defined, consult the remark following Definition 1.2. Then  $F(f, g)$  is defined and not zero only if  $g = l\phi_i$  for some  $l$  and  $i$  and then only for  $f$ 's in the form  $k\phi_i$ . For each such  $g$  it is obviously linear. A similar statement holds for  $f$ .

Now the domain of  $F$  is completely linear. For let  $f \otimes g$  be such that  $f \otimes g = \sum_{i=1}^n \lambda_i \phi_{n_i} \otimes \phi_{n_i}$ . As in §1, Lemma 1.1 above, we see that  $f = \sum_{i=1}^n c_i \phi_{n_i}$ ,  $g = \sum_{i=1}^n d_i \phi_{n_i}$ . Hence

$$\begin{aligned} \sum_{i=1}^n \lambda_i \phi_{n_i} \otimes \phi_{n_i} &= f \otimes g = \left( \sum_{i=1}^n c_i \phi_{n_i} \right) \otimes \left( \sum_{j=1}^n d_j \phi_{n_j} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^n c_i d_j \phi_{n_i} \otimes \phi_{n_j}. \end{aligned}$$

Since the  $\phi_{n_i} \otimes \phi_{n_j}$  are mutually orthogonal, it follows that the matrix  $(\lambda_i \delta_{i,j})$ ,  $(i, j = 1, \dots, n)$ , must equal the matrix  $(c_i d_j)$ ,  $(i, j = 1, \dots, n)$ . Since the latter matrix is of rank at most one, the former is also, which means that at most one of  $\lambda_i$ 's is not zero. Hence  $f \otimes g = 0$  or  $f \otimes g = \lambda_k \phi_{n_k} \otimes \phi_{n_k}$ , and in either case  $f \otimes g$  is in the domain of  $F$ .

Since  $\phi_1 \otimes \phi_1$  and  $\phi_2 \otimes \phi_2$  are in the domain of  $F$  but  $\phi_1 \otimes \phi_2$  is not,  $F$  does not have a rectangular domain.

EXAMPLE 2. We give an example of a bilinear transformation which is not completely linear. Let  $\phi_1, \phi_2, \phi_3, \phi_4$  be four orthonormal elements. Let, for  $\tau \neq 0$ ,

$$F(k(\phi_1 + \tau\phi_2), l(\phi_1 - (1/\tau)\phi_2)) = kl(\frac{1}{2}(1 + \tau)\phi_1 + \frac{1}{2}(1 - \tau)\phi_4)$$

and also

$$F(k\phi_1, l\phi_2) = kl\phi_2, \quad F(k\phi_2, l\phi_1) = kl\phi_3.$$

The proof that  $F$  is bilinear is similar to that given for Example 1.

If we take  $\tau = 1$  and  $\tau = -1$ , then we have, respectively,

$$F(\phi_1 + \phi_2, \phi_1 - \phi_2) = \phi_1, \quad F(\phi_1 - \phi_2, \phi_1 + \phi_2) = \phi_4.$$

We also notice that

$$(\phi_1 + \phi_2) \otimes (\phi_1 - \phi_2) = (\phi_1 - \phi_2) \otimes (\phi_1 + \phi_2) + 2\phi_2 \otimes \phi_1 - 2\phi_1 \otimes \phi_2.$$

But

$$\begin{aligned} F(\phi_1 + \phi_2, \phi_1 - \phi_2) &= \phi_1 \neq \phi_4 + 2\phi_3 - 2\phi_2 \\ &= F(\phi_1 - \phi_2, \phi_1 + \phi_2) + 2F(\phi_2, \phi_1) - 2F(\phi_1, \phi_2). \end{aligned}$$

It should also be noticed, however, that the domain of  $F$  is completely

linear. For if  $f \otimes g \neq 0$  is a linear combination of elements of the domain of  $F$ , one readily sees that it must be in the form

$$f \otimes g = \lambda(\phi_1 \otimes \phi_1 - \phi_2 \otimes \phi_2) + \mu\phi_1 \otimes \phi_2 + \nu\phi_2 \otimes \phi_1$$

for some constants  $\lambda$ ,  $\mu$ , and  $\nu$ . As in the argument given in Example 1, this implies that the determinant  $-\lambda^2 - \mu\nu = 0$ . Now if  $\lambda$  is zero, either  $\mu$  or  $\nu$  is zero and  $f \otimes g = \nu\phi_2 \otimes \phi_1$  or  $f \otimes g = \mu\phi_1 \otimes \phi_2$ ; hence  $f \otimes g$  is in the domain of  $F$ . If, however,  $\lambda$  is not zero, it may be taken as 1, and then  $\mu\nu = -1$ . Now if we let  $\tau = \mu$ , then  $\nu = -1/\tau$  and

$$f \otimes g = (\phi_1 + \tau\phi_2) \otimes (\phi_1 - (1/\tau)\phi_2).$$

This also is in the domain of  $F$ .

EXAMPLE 3. We give an example of an  $F$  and  $\bar{F}$ , related as in Definition 5.2, in which  $F$  is closed with a rectangular or even a rectangular symmetric domain for which  $\bar{F}$  is nevertheless a proper extension of  $F$ .

We begin as follows. Let  $\{\phi_i\}$ ,  $(i=0, 1, 2, \dots)$ ,  $\{\psi_j\}$ ,  $(j=0, 1, 2, \dots)$ ,  $\{\chi_{i,j}\}$ ,  $(i, j=1, 2, \dots)$ , be three infinite orthonormal sets of elements. Let a transformation  $T''$  from  $\mathfrak{H} \otimes \mathfrak{H}$  to  $\mathfrak{H}$  be defined by the equations

$$(1) \quad T''(\phi_0 + (1/n)\phi_n) \otimes (\psi_0 + (1/m)\psi_m) = ((n-m)^2 + 1)\chi_{n,m},$$

for  $n, m=1, 2, \dots$ .

Let  $T'$  be the least linear extension of  $T''$ . (The existence of  $T'$  is easily demonstrated since the

$$(\phi_0 + (1/n)\phi_n) \otimes (\psi_0 + (1/m)\psi_m)$$

are linearly independent.) Let  $F'$  be the bilinear transformation associated with  $T'$  as in Theorem 5 above. Let  $F$  be the closure of  $F'$ .

We show that  $T'$  is closable. Let  $\omega_1, \omega_2, \dots$  be a sequence in the domain of  $T'$ , such that  $\omega_i \rightarrow \omega$  and  $\sigma_i = T'\omega_i \rightarrow \sigma \neq 0$ . Then

$$(2) \quad \sigma = \sum_{i,j} s_{i,j} \chi_{i,j},$$

where for some pair  $n$  and  $m$ ,  $s_{n,m} \neq 0$ . Now since  $\omega_k$  is in the domain of  $T'$ ,

$$(3) \quad \omega_k = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} p_{i,j}^{(k)} (\phi_0 + (1/i)\phi_i) \otimes (\psi_0 + (1/j)\psi_j)$$

and

$$\sigma_k = T'\omega_k = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} p_{i,j}^{(k)} ((i-j)^2 + 1) \chi_{i,j}.$$

Now since  $\sigma = \lim_{k \rightarrow \infty} \sigma_k$ ,

$$s_{n,m} = (\sigma, \chi_{n,m}) = \lim_{k \rightarrow \infty} (\sigma_k, \chi_{n,m}) = \lim_{k \rightarrow \infty} p_{n,m}^{(k)}((n-m)^2 + 1).$$

Thus

$$(4) \quad \lim_{k \rightarrow \infty} p_{n,m}^{(k)} = \frac{s_{n,m}}{(n-m)^2 + 1} \neq 0.$$

On the other hand,

$$((\phi_0 + (1/i)\phi_i) \otimes (\psi_0 + (1/j)\psi_j), \phi_n \otimes \psi_m) = (1/n \cdot m) \delta_n^i \delta_m^j$$

for  $n$  and  $m \geq 1$ . With (3), this implies

$$(\omega_k, \phi_n \otimes \phi_m) = (1/n \cdot m) p_{n,m}^{(k)}.$$

Since  $\omega = \lim_{k \rightarrow \infty} \omega_k$ , (4) now implies  $(\omega, \phi_n \otimes \phi_m) \neq 0$ . Hence  $\omega \neq 0$ . Thus if  $\{\omega, \sigma\}$  is a pair in the closure of the graph of  $T'$ , then  $\sigma \neq 0$  implies  $\omega \neq 0$ . Hence closure is the graph of a transformation or  $[T']$  exists.

Let  $\bar{F}$  be related to  $[T']$  as in Theorem 5.

We now wish to discuss  $F$ , the closure of  $F'$ , and, in particular, its domain. To do this, we define  $F_0$  as the bilinear transformation whose domain consists of pairs  $f \otimes g$  in the form

$$f = \sum_{i=1}^{\infty} a_i(\phi_0 + (1/i)\phi_i), \quad g = \sum_{j=1}^{\infty} b_j(\psi_0 + (1/j)\psi_j),$$

and for which

$$(5) \quad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |a_i|^2 \cdot |b_j|^2 ((i-j)^2 + 1)^2 < \infty,$$

and which is defined by the equation

$$(6) \quad F_0(f, g) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_i b_j ((i-j)^2 + 1) \chi_{i,j}.$$

If either  $f$  or  $g$  is zero, we simply demand that the other be in the manifold determined by the  $\psi$ 's or  $\phi$ 's, respectively. It is easily seen that for  $f$  and  $g$ , both not zero (5) is equivalent to

$$(7) \quad \sum_{i=1}^{\infty} |a_i|^2 i^4 < \infty, \quad \sum_{j=1}^{\infty} |b_j|^2 j^4 < \infty.$$

These last conditions also insure that for  $f$  and  $g$  both not zero  $f$  shall be in the form  $\sum_{i=1}^{\infty} a_i(\phi_0 + (1/i)\phi_i)$  since for this it is sufficient that  $\sum_{i=1}^{\infty} a_i$  exist and that  $\sum_{i=1}^{\infty} |a_i|^2/i^2 < \infty$ . The second condition is trivial, and the first follows from the fact that

$$\sum_{i=1}^n |a_i| = \sum_{i=1}^n |a_i| i^2 (1/i^2) \leq \left( \sum_{i=1}^n |a_i|^2 i^4 \right)^{1/2} \left( \sum_{i=1}^n 1/i^4 \right)^{1/2}.$$

Now  $F_0$  is closed. For suppose a sequence  $f_n \otimes g_n \rightarrow f \otimes g$  for which  $F_0(f_n, g_n) \rightarrow h$ . If  $g=0$ , then  $f_n \otimes g_n \rightarrow 0$ , and by a proper choice of  $\lambda_n$ , we can let  $f'_n = \lambda_n f_n$ ,  $g = (1/\lambda_n)g_n$  in such a way that while  $F_0(f'_n, g'_n) = F_0(f_n, g_n) \rightarrow h$ , we also have both  $f_n \rightarrow 0$  and  $g'_n \rightarrow 0$ . Thence  $a_i^{(n)} \rightarrow 0$ ,  $b_i^{(n)} \rightarrow 0$ , and (6) then implies that  $(h, \chi_{i,j}) = 0$  for every  $i$  and  $j$  and hence that  $h=0$ . But obviously  $F_0(f, 0) = 0$ . A similar argument will apply if  $f=0$ .

Now if neither  $f$  nor  $g$  is 0, then we can find sequences  $f'_n$  and  $g'_n$  such that  $f'_n \rightarrow f$ ,  $g'_n \rightarrow g$  with  $F_0(f'_n, g'_n) = F_0(f_n, g_n) \rightarrow h$ . These results and (6) for  $f'_n$  and  $g'_n$  yield that

$$(h, \chi_{i,j}) = a_i b_j ((i-j)^2 + 1).$$

Hence (5) holds for  $f \otimes g$ , and, furthermore, when we form  $F_0(f, g)$  by (6), we get  $F_0(f, g) = h$ .

Now  $F_0$  is obviously an extension of  $F'$ , and since it is closed it must be an extension of the closure of  $F'$ ; that is,  $F$ . But, on the other hand, if  $F_0(f, g) = h$ , then by taking partial sums in the expressions for  $f$  and  $g$ , given above, we see that  $\{f \otimes g, h\}$  is the limit of the pairs  $\{f_n \otimes g_n, h_n\}$ , where  $F'(f_n, g_n) = h_n$ . This implies that  $F_0$  is included in  $F$ . Thus we have  $F = F_0$ .

Now condition (7) above implies that the domain of  $F$  is rectangular (Definition 1.5); hence, by Theorem 6,  $F$  is completely linear. Now let  $Tf \otimes g = F(f, g)$  as in Theorem 4. Furthermore  $F = F_0$  is closed. If  $\phi_i = \psi_i$ , the domain is even symmetric.

To show our statement then, it remains only to prove that (a)  $F$  and  $\bar{F}$  are related as in Definition 5.2 and (b)  $\bar{F}$  is a proper extension of  $F$ .

(a) follows from the fact that since  $F$  is the closure of  $F'$ ,  $[l(\mathfrak{F}')] = [l(\mathfrak{F})]$  and hence  $[T] = [T']$ .

To prove (b), we note first that  $\phi_0 \otimes \psi_0 = \bar{f}$  is not in the domain of  $F = F_0$ . For otherwise we would have that

$$\phi_0 = \left( \sum_{i=1}^{\infty} a_i \right) \phi_0 + \sum_{i=1}^{\infty} (a_i/i) \phi_i$$

or  $a_i = 0$  for every  $i$  and  $\sum_{i=1}^{\infty} a_i = 1$ .

But, on the other hand, assume

$$\bar{f}_n = \sum_{i=1}^n (1/n)(\phi_0 + (1/i)\phi_i) \otimes (\psi_0 + (1/i)\psi_i) \in \mathfrak{S} \otimes \mathfrak{S}.$$

Then we obtain the relation

$$\begin{aligned}
\|\bar{f}_n - \bar{f}\|^2 &= \left\| \sum_{i=1}^n (1/n)(\phi_0 + (1/i)\phi_i) \otimes (\psi_0 + (1/i)\psi_i) - \phi_0 \otimes \psi_0 \right\|^2 \\
&= \left\| \sum_{i=1}^n ((1/n i)\phi_i \otimes \psi_0 + (1/n i)\phi_0 \otimes \psi_i + (1/i^2 n)\phi_i \otimes \psi_i) \right\|^2 \\
&= \sum_{i=1}^n (2/n^2 i^2 + 1/i^4 n^2) = (1/n^2) \left( \sum_{i=1}^n 2/i^2 + 1/i^4 \right) \\
&\rightarrow 0.
\end{aligned}$$

Also by (1) above, we obtain

$$\|T'f_n\|^2 = \left\| \sum_{i=1}^n (1/n)\chi_{i,i} \right\|^2 = \sum_{i=1}^n (1/n^2) = 1/n = 0.$$

Hence  $f_n \rightarrow \phi_0 \otimes \psi_0$ ,  $T'f_n \rightarrow 0$ . Thus since  $T'$  has a closure, we have  $[T]\phi_0 \otimes \psi_0 = [T']\phi_0 \otimes \psi_0 = 0$ . This implies that  $\bar{F}(\phi_0, \psi_0)$  exists. Since  $F(\phi_0, \psi_0)$  does not, we have shown (b).

EXAMPLE 4. Let  $F(f, g) = f, Jg)h$ , where  $(,)$  denotes the inner product,  $J$  is a conjugation (cf. [14], pp. 357-365),  $h$  is some fixed element, and  $f$  and  $g$  are any two elements. Then  $F$  is easily seen to be bounded, bilinear, and with domain all  $f \otimes g$ . Thus  $F$  is completely linear by Theorem 6. Let  $T$  be related to  $F$ , as in Theorem 4.

But  $F$  is not hyperclosable. For if we let  $\phi_1, \phi_2, \dots$  be an infinite orthonormal set, then if  $f_n = \sum_{i=1}^n (1/n)\phi_i \otimes J\phi_i$ ,  $Tf_n = h$  and  $f_n \rightarrow 0$ . Thus  $T$  has no closed extension.

#### REFERENCES

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.
2. ———, *Studia Mathematica*, vol. 8 (1938), pp. 36-44.
3. F. L. Hitchcock, *Journal of Mathematics and Physics*, vol. 8 (1929), p. 83. (This memoir contains references to preceding work on the finite dimensional case.)
4. M. Kerner, *Annals of Mathematics*, (2), vol. 41 (1937), pp. 208-248.
5. S. Mazur and W. Orlicz, *Studia Mathematica*, vol. 5 (1935), pp. 50-68, 179-189.
6. F. J. Murray, *these Transactions*, vol. 37 (1935), pp. 301-338.
7. F. J. Murray and J. von Neumann, *Annals of Mathematics*, (2), vol. 37 (1936), pp. 116-229.
8. ———, *these Transactions*, vol. 41 (1937), pp. 208-248.
9. J. von Neumann, *Mathematische Annalen*, vol. 102 (1929-1930), pp. 370-427.
10. ———, *Annals of Mathematics*, (2), vol. 32 (1931), pp. 191-226.
11. ———, *Annals of Mathematics*, vol. 33 (1932), pp. 294-310.
12. R. Oldenburger, *Annals of Mathematics*, (2), vol. 35 (1934), pp. 622-657 (two papers).
13. ———, *these Transactions*, vol. 39 (1936), pp. 422-455.
14. M. H. Stone, *Linear Transformations in Hilbert Space*, American Mathematical Society Colloquium Lectures, vol. 15, New York, 1932.

COLUMBIA UNIVERSITY,  
NEW YORK, N. Y.

26

111

